

# Predicting Uterine Deformation Due to Applicator Insertion in Pre-Brachytherapy MRI Using Deep Learning

by

Shrimanti Ghosh

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Computing Science

University of Alberta

© Shrimanti Ghosh, 2023

# Abstract

In locally advanced cervical cancer (LACC), brachytherapy (BT) remains the gold standard for boosting to curative doses in radiotherapy. Progress towards balancing target and routine tissue dosimetry for better clinical outcomes has been made possible by magnetic resonance imaging (MRI)-guided imaging data for treatment planning and by improving traditional BT applicator geometry. However, further evolving the original “one size fits all” approach towards truly personalized BT delivery requires improvements along the entire BT care pathway. From the pre-insertion applicator selection to the post-insertion treatment plan optimization, many factors need to be optimized for each patient for the best possible tumour and morbidity outcomes. The expanding collection and sharing of data, increased computational power in machine learning (ML) and artificial intelligence (AI), deep learning (DL) are rapidly transforming society, and offer the potential for similar transformation within health care. The success of these algorithms is founded on their judicious capability for detecting complex patterns even in heterogeneous datasets.

This thesis aims to develop ML and DL-based models and free-form deformation methods for building predictive models to guide BT processes and decisions for consistently better personalization in LACC. First, an automated segmentation algorithm is proposed to delineate the uterus from the background on MRI using a deep convolutional neural network (CNN) architecture (Inception-V4) along with auto-encoders. After automated uterus segmenta-

tion, a modified version of another deep CNN model i.e. U-net is utilized to predict the at-BT uterus shape from pre-BT MRI. Finally, a shape-based non-rigid registration (free form deformation) algorithm is proposed to measure or quantify the amount of complex and large deformations of the uterus structure due to BT applicator insertion. *The study deals with the very challenging and complex problem of predicting the large anatomical deformations from pelvic MR-images due to BT applicator insertion.* The proposed method achieved an average Dice Coefficient of 94.8% and a Hausdorff distance of 3.06 mm, whereas the U-net yielded 92.4% and 6.7 mm for the Dice score and Hausdorff Distance metrics, respectively in the uterus segmentation task. The quantitative evaluations demonstrated that the proposed method performed significantly better than U-net in terms of both Dice Coefficient and Hausdorff Distance. After that, a pre-trained modified U-net is proposed to predict the at-BT uterus position from only the pre-BT MRI. This method yielded an average Dice score of 89.5% and a Hausdorff distance of 3.6 mm in predicting the uterine deformation automatically. Large anatomy deformations before and at the time of BT insertion were observed for most patients due to the insertion of the BT applicator. In order to quantify this deformation, a free-form deformation model-based non-rigid registration method is proposed. The applicator’s presence introduces a median uterine surface point-to-point displacement of 25.0 [10.0 - 62.5] mm and a median uterine cavity point-to-point displacement of 40.0 [12.0 - 68.0] mm from the pre-BT position.

The challenge in implementing this algorithm was the inter-patient anatomical dissimilarity and extreme intra-patient uterine deformation from pre-BT to at-BT in the dataset. Increasing the size of our training dataset, with the inclusion of more heterogeneous images with anatomical variability, will improve the prediction accuracy of this DL-based algorithm. Our proposed

DL-based model, despite being trained on heterogeneous and complex deformations, can successfully predict uterine distortion automatically due to applicator insertion using only the pre-BT MRI, which can guide the clinicians in selecting the most suitable applicator component and configuration ahead of the actual insertion procedure. These promise better, faster, and more streamlined clinical/technical decision-making before BT applicator insertion and plan optimization, potentially enabling more consistent application of BT personalization for LACC and improved dosimetric outcomes.

# Preface

The thesis is an original work by Shrimanti Ghosh, Ph.D. candidate in the Department of Computing Science, University of Alberta. This is a collaborative project between the Department of Computing Science and the Department of Oncology of the University of Alberta. This research has been funded by the Alberta Women’s Health Foundation through the Women and Children’s Health Research Institute.

During my Ph.D. study, I have received University of Alberta Doctoral Recruitment Scholarship (CAD 10000) in 2018, Jeffrey R Sampson Memorial Graduate Prize (CAD 1600) by the Department of Computing Science, University of Alberta, for achieving the highest GPA (4.0 / 4.0) during Ph.D. in 2019, Graduate Student Teaching Award for excellent performance as a Teaching Assistant in 2020, Verna Tate Graduate Scholarship in Science (CAD 10000) for superior academic achievements in 2020, J Gordin Kaplan Graduate Student Award (CAD 1500) for academic and scholarly achievements in 2022.

Our proposed novel method on automated organ segmentation using deep convolutional neural networks along with autoencoders was published at *2020 IEEE International Symposium on Biomedical Imaging* in Iowa, USA.

The initial results with 80 patients were published at *2021 Canadian Organization of Medical Physics* and *Women and Children’s Health Research Institute; WCHRI Research Day* in 2021.

Results with 92 patients along with some modified CNN models was published in *2022 IEEE Engineering in Medicine and Biology Society* in Glasgow, UK.

I performed for all the major work as the primary author here and was responsible for implementation, experimentation and analysis.

*To my whole family,  
For their infinite love and support at each step of my life ...*

*All our dreams can come true, if we have the courage to pursue them.*

– Walt Disney.

# Acknowledgements

I am deeply indebted to my respected supervisors Dr. Pierre Boulanger, Dr. Geetha Menon, and Dr. Kumaradevan Punithakumar for their continuous support and valuable guidance which helped me to complete my Ph.D. project. Their immense knowledge and plentiful experience have encouraged me in all the time of my academic research career.

I am grateful to Dr. Fleur Huang and Dr. Geetha Menon for providing the manual/ground truth segmentation for our experiments as this task was extremely rigorous and time consuming.

I am also thankful to my fellow labmates, colleagues, my office mates and friends for their honest cooperation in my necessity. Above all, I wish to thank our University of Alberta, for building up my career and creating a passion for teaching and research in me which would also help me in developing my professional career. Our University and our Department of Computing Science really encouraged me and inspired me in every way.

Finally, I would like to express my gratitude to my family, without their tremendous support and encouragement, it would be impossible for me to complete my study. Their belief in me has kept my motivation high during this process.

# Table of Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Motivation . . . . .	7
1.2	Challenges . . . . .	10
1.3	Thesis Objectives . . . . .	13
1.4	Thesis Methodology . . . . .	14
1.4.1	Patient Information and Imaging Dataset . . . . .	14
1.4.2	Manual Segmentation/Ground Truth . . . . .	14
1.5	Clinical relevance . . . . .	15
1.6	Thesis Contributions . . . . .	16
1.6.1	Contributions . . . . .	16
1.6.2	Papers Published in Peered Reviewed Conferences . . . . .	18
<b>2</b>	<b>Background</b>	<b>19</b>
2.1	Application of Artificial Intelligence in Brachytherapy . . . . .	19
2.1.1	AI in Brachytherapy Pre-Planning . . . . .	20
2.1.2	AI in assisting procedure in BT operating room . . . . .	21
2.2	Organ Delineation Using Deep Learning . . . . .	22
2.2.1	Automatic Segmentation in Cervical Cancer using Deep Learning . . . . .	23
2.3	Applicator Segmentation and Reconstruction . . . . .	25
2.4	Dealing with Large Deformations in Cervical Cancer BT Imaging . . . . .	27
2.5	Various Non-rigid Image Registration . . . . .	28
2.6	Diffeomorphic Image Registration . . . . .	30
2.7	Modeling Soft Tissue Deformations . . . . .	32
2.7.1	Physical methods to simulate deformable materials . . . . .	36
2.7.2	Mass Spring Model (MSM) . . . . .	37
2.7.3	Application of MSM in Brachytherapy . . . . .	38
2.8	Limitations of Various DIR Algorithms in Brachytherapy . . . . .	39

<b>3</b>	<b>Automated Intensity-based Image Registration</b>	<b>41</b>
3.1	Overview . . . . .	41
3.2	Image Registration Methodology . . . . .	42
3.3	Registration of Brachytherapy MRIs . . . . .	46
3.4	Our Methodology . . . . .	47
3.4.1	Intensity-based Image Registration . . . . .	47
3.4.2	Affine Transformation . . . . .	47
3.4.3	Similarity Metric . . . . .	49
3.4.4	Optimizer . . . . .	51
3.5	Comparison with Other Methods . . . . .	53
3.5.1	Optical Flow . . . . .	53
3.5.2	Elastix . . . . .	54
3.6	Results . . . . .	54
3.7	Discussion . . . . .	58
<b>4</b>	<b>Automated Uterus Segmentation Using CNN</b>	<b>59</b>
4.1	Overview . . . . .	59
4.2	Various Medical Image Segmentation Methods . . . . .	60
4.3	Uterus Segmentation using Deep CNN Inception-V4 . . . . .	62
4.3.1	Proposed Deep ConvNet Architecture . . . . .	62
4.3.2	Utilizing Transfer Learning . . . . .	62
4.3.3	Dimensionality Reduction Using Autoencoders . . . . .	63
4.4	Uterus Segmentation using Deep U-Net . . . . .	66
4.4.1	U-net Architecture . . . . .	66
4.5	Results . . . . .	69
4.5.1	Data Preprocessing . . . . .	69
4.5.2	Data Augmentation . . . . .	70
4.5.3	Implementation Details . . . . .	71
4.5.4	Quantitative Evaluation Metrics . . . . .	71
4.5.5	Uterus Segmentation Results - Our Proposed Method vs U-Net . . . . .	72
4.6	Discussion . . . . .	72
<b>5</b>	<b>Uterus Structure Prediction Using Deep Learning</b>	<b>78</b>
5.1	Overview . . . . .	78
5.2	U-Net Architecture . . . . .	81
5.3	Fine-tuning details of proposed U-net architecture . . . . .	82
5.3.1	Backbone U-Net architecture . . . . .	82
5.3.2	Bidirectional Feature Network . . . . .	82
5.3.3	ReLU Activation Functions . . . . .	84
5.3.4	Data augmentation . . . . .	85

5.4	Comparison with U-Net++ Architecture . . . . .	86
5.5	Results . . . . .	87
5.5.1	Implementation Details . . . . .	87
5.5.2	Automated Uterus Structure Prediction Using CNN . . . . .	88
5.6	Discussion . . . . .	89
<b>6</b>	<b>Non-rigid registration/Free form deformation</b>	<b>94</b>
6.1	Overview . . . . .	94
6.2	Methodology . . . . .	96
6.2.1	Registration Error (Squared Distance) . . . . .	97
6.2.2	Free Form Deformation . . . . .	99
6.2.3	Registration Error (Squared Distance) . . . . .	99
6.3	Results . . . . .	101
6.4	Discussion . . . . .	102
<b>7</b>	<b>Conclusion and Future Work</b>	<b>108</b>
7.1	Discussion . . . . .	108
7.1.1	Application of Mass-Spring Methods (MSMs) . . . . .	108
7.1.2	Application of Generative Adversarial Networks . . . . .	115
7.2	Conclusion . . . . .	120
7.2.1	To Develop User Friendly Software/Library . . . . .	122
7.3	Future Work . . . . .	124
	<b>Bibliography</b>	<b>126</b>

# List of Tables

3.1	Evaluation of Intensity-based image registration method. <b>Structural Similarity Index Measure (SSIM)</b> between two images are reported. SSIM between fixed image and registered image, fixed image and moving image are reported. A value closer to 1 indicates better image similarity. . . . .	54
3.2	Evaluation of Intensity-based image registration method. <b>Mutual Information (MI)</b> between two images are reported. MI between fixed image and registered image, fixed image and moving image are reported. The larger the value, the greater the relationship between the two images. . . . .	55
4.1	Evaluation of automated segmentation results by the proposed method and U-net in terms of Dice Coefficient (DC) and Hausdorff Distance (HD) in comparison to expert manual segmentation. The Mean $\pm$ Standard Deviation values are reported for each metric. The higher the Dice coefficient or the lower the Hausdorff distance the better the results. . . . .	72
5.1	Evaluation of uterus prediction results by the proposed method in terms of Dice Coefficient (DC) and Hausdorff Distance (HD) in comparison to expert manual segmentation. The Mean $\pm$ Standard Deviation values are reported for each metric. The higher the Dice coefficient or the lower the Hausdorff distance the better the results. . . . .	88
6.1	Quantitative analysis of anatomical deformation (uterus, uterine canal, vaginal canal) due to applicator insertion between pre-BT and at-BT anatomical structures. The Mean $\pm$ Standard Deviation values of the Mean Squared Distance are reported here.	101

# List of Figures

1.1	This figure shows the shrinkage of the initial tumor (a: before EBRT) following EBRT (b: after EBRT). Note how the tumor has shrunk. . . . .	2
1.2	Figure shows two commonly used Intracavitary/Interstitial IC/IS applicators: (a) tandem and ovoid and (b) tandem and ring, both with IS option (Elekta, Stockholm, Sweden). (c) and (d) show the placement of these two applicators in patients, respectively. . . . .	4
1.3	MRIs taken (a) prior to BT (pre-BT) applicator (intrauterine tandem and intravaginal ring; middle) insertion and (b) large/complex uterine deformation post-applicator insertion (at-BT). . . . .	5
1.4	Some unusual and challenging datasets in our study where the uterus uniquely changes its position and shape . . . . .	11
1.5	Some unusual and challenging cases in our study where the size of the uterus varies greatly from patient to patient and showing the diversity in our dataset . . . . .	12
1.6	T2 weighted sagittal MR images of the same patient taken at (a) pre-BT and (b) at-BT. Both images show the 4 regions of interest used in this study. Green - uterus, yellow - uterine canal, blue - vaginal canal, and red - external os. . . . .	15
3.1	Our goal is to align the bone structure properly between pre-BT (moving) and at-BT (fixed) MRI images . . . . .	46
3.2	Automated image registration algorithm to align spine between pre-BT and at-BT MRI . . . . .	52
3.3	To align the bone structure, the moving image is transformed vertically and horizontally to align with the fixed image . . . . .	56
3.4	Another example of the registration method where the organs change shape and position drastically . . . . .	56
3.5	There is a large uterine deformation in this example which makes the registration problem harder to solve as the only point of reference is the bony anatomy . . . . .	56
3.6	Another example of the registration method where the uterus is retroverted . . . . .	57
4.1	The schema for the pure Inception-V4 network . . . . .	65

4.2	Components of proposed neural net approach which consists of fine-tuned CNN architecture combining decoder part of auto-encoder . . . . .	65
4.3	Illustration of U-Net convolution network structure. The left side of the U-shape is the encoding stage, also called the contraction path and the right side of the U-shape, also called the expansion part, consists of the decoding stage and the up-sampling process . . . . .	67
4.4	Illustration of U-Net convolution network structure for our problem formulation . . . . .	68
4.5	Transformation from DICOM to patient-coordinate system . .	69
4.6	Loss - (MSE) vs. Epoch curve for training and validation for CNN with autoencoder . . . . .	73
4.7	The automated segmentation results of CNN with autoencoders and U-net (in square brackets) with different Dice score (DC) values. The red, green, and yellow contours represent the ground truth, segmentation by CNN with autoencoders and U-net, respectively. . . . .	74
4.8	Unusual results predicted by DL methods. Segmentation results of CNN with autoencoders and U-net (in square brackets) with different Dice score (DC) values. The red, green, and yellow contours represent the ground truth, segmentation by CNN with autoencoders and U-net, respectively. . . . .	75
5.1	Architecture of the proposed U-net model for predicting the deformed at-BT uterus. . . . .	85
5.2	Illustration of U-Net++ Architecture . . . . .	87
5.3	Loss - (MSE) vs. Epoch curve for training and validation for U-net . . . . .	89
5.4	Results with different Dice score values for predicting the uterus structure using U-net (green) and U-Net++ (yellow). The red contour represents the manual ground truth. The corresponding input MRI (pre-BT) is also displayed in the right to have a clear idea about the problem formulation and the output results. . .	90
5.5	Results with different Dice score values for predicting the uterus structure using U-net (green) and U-Net++ (yellow). The red contour represents the manual ground truth. The corresponding input MRI (pre-BT) is also displayed in the right to have a clear idea about the problem formulation and the output results. . .	91
5.6	Results with different Dice score values for predicting the uterus structure using U-net (green) and manual ground truth (red). As for any test cases, there is no at-BT image, that is why . .	92

6.1	Registering the <b>pre-BT uterine canal</b> with the applicator shape. Red - applicator shape and Blue - pre-BT uterine canal. The iteration number and the distance between pre-BT and at-BT shapes are given here. . . . .	103
6.2	Registering the <b>pre-BT vaginal canal</b> with the applicator shape. Red - applicator shape and Blue - pre-BT vaginal canal. The iteration number and the distance between pre-BT and at-BT shapes are given here. . . . .	104
6.3	Registering the <b>pre-BT uterus</b> with at-BT uterus. Red - at-BT uterus and Blue - pre-BT uterus. The iteration number and the distance between pre-BT and at-BT uterus shapes are given here. . . . .	105
6.4	Registering the <b>pre-BT uterus</b> with at-BT uterus. Red - at-BT uterus and Blue - pre-BT uterus. The iteration number and the distance between pre-BT and at-BT uterus shapes are given here. . . . .	106
7.1	Results of uterus deformation using Mass-Spring Model where yellow structure is the uterus and red structure is the applicator model . . . . .	111
7.2	Results of uterus deformation using Mass-Spring Model in presence of the rigid bone structure . . . . .	112
7.3	Results of CycleGAN in predicting at-BT MRI from pre-BT MRI	119

purposes.

# Chapter 1

## Introduction

With an estimated worldwide incidence of 570,000 cases and 311,000 deaths in 2018, cervical cancer ranks as the fourth most frequently diagnosed cancer and the fourth leading cause of cancer death in women [1]. It is projected that 1,450 cases of cervical cancer will be diagnosed in Canada in 2022, making it the 15<sup>th</sup> most commonly diagnosed cancer among females in Canada [2]. More than 70% of Stage 1 cervical cancers diagnosed in females are in the age group between 18–39 years [1]. Globally, cervical cancer management is still a challenge. Despite the availability of screening, vaccination, and cutting-edge treatment technologies, the survival rate from this disease is poor in many parts of the world. The first line of treatment for cervical cancer is surgery [3]. In cases where surgery is not possible, the treatment for locally advanced cervical cancer (LACC) involves a multimodal approach that includes a combination of external beam radiotherapy (EBRT) with concurrent chemotherapy followed by brachytherapy (BT) [4, 5]. The EBRT treatment is applied to the pelvic lymph nodes, parametria, and primary tumour, to a dose adequate to control the microscopic disease. EBRT aims x-rays (radiation source) at the cancer from a machine outside the body, When EBRT is

used as the main treatment for cervical cancer, it is usually combined with chemotherapy. Brachytherapy (BT) is a form of radiation therapy where the radiation source is placed inside or next to the tumour. BT allows doctors to deliver higher doses of radiation to more specific areas of the body than conventional radiation therapy (EBRT) projects radiation. The addition of BT boosts the gross tumour and improves disease control and survival [6, 7]. Figure 1.1 shows the initial tumor size and shrinkage of the tumor after EBRT treatment.



(a) Tumor (in red) before EBRT

(b) Tumor (in red) after EBRT

Figure 1.1: This figure shows the shrinkage of the initial tumor (a: before EBRT) following EBRT (b: after EBRT). Note how the tumor has shrunk.

After concurrent EBRT and chemotherapy, a focused volume-prescribed radiation dose is delivered to the rest of the tumour in the cervix using BT. This is an essential component of LACC treatment and plays a critical role. Multiple studies have demonstrated a decrease in local recurrence and improved patient outcomes when BT is used [8–10]. The success of BT requires extreme conformity, with the delivery of a high radiation dose directly to the tumour while sparing surrounding normal tissues due to the rapid radiation dose fall-off beyond the implanted tumour volume. BT involves the placement

of a radioactive source very close to or inside the tumour to help deliver a high radiation dose to the tumour. In North America, BT for cervical cancer consists of two approaches: high-dose-rate (HDR) treatments delivered daily or weekly) or pulsed-dose-rate (PDR), a low-dose hourly treatment delivered over a few days. Both treatments appear to be equivalent in terms of survival outcomes based on existing retrospective and prospective studies [11,12]. BT for cervical cancer can be performed using intracavitary (IC), interstitial (IS), or a combination technique (IC/IS). Figure 1.2 (a and b) shows the two most commonly used IC/IS applicators. The vaginal component, which mainly includes the ovoids (Figure 1.2 (a and c)) and ring (Figure 1.2 (b and d)), sits in the upper part of the vagina, i.e., close to the cervix, and the tandem passes through the internal orifice of the cervix uteri (os) into the intrauterine canal (Figure 1.2 (c and d)). The radioactive source travels through these applicators to treat the upper vagina, cervix, and uterus [4]. IS catheters (small tubes) are placed in and around the residual disease, anchored using the vaginal component, to achieve better lateral dose coverage [3]. The choice of the applicator and the technique depends primarily on the disease extent and patient anatomy. A treatment strategy is based on information obtained from imaging, mainly magnetic resonance imaging (MRI), acquired post-EBRT [13, 14].

One of the technical advancements in cervical cancer treatment involves using MRI guidance for better visualization of the tumour and surrounding critical structures. This leads to better dose optimization and could improve treatment outcomes. Two MRI scans are performed: one close to or at the end of EBRT (**pre-BT**) and one after BT applicator insertion (**at-BT**) as shown in Figure 1.3. Before selecting and inserting an applicator, the physicians make several educated predictions, mainly based on the pre-BT MR images,

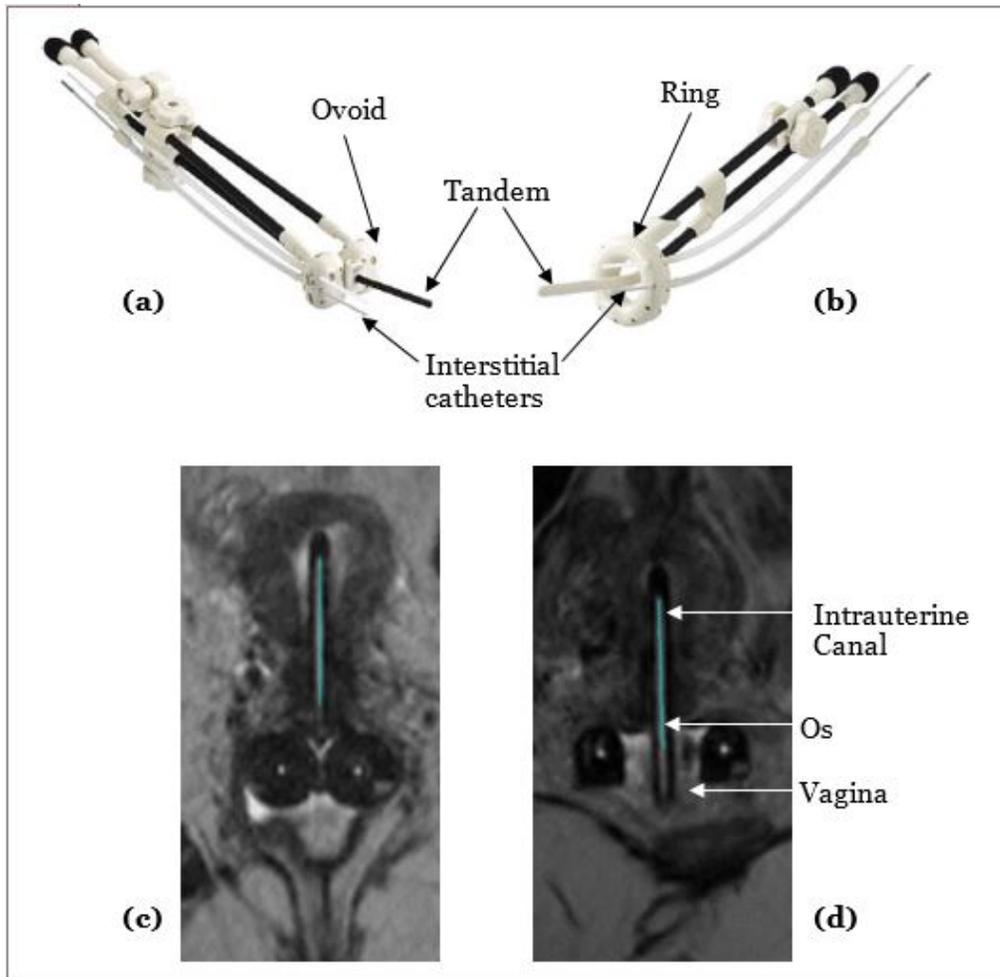


Figure 1.2: Figure shows two commonly used Intracavitary/Interstitial IC/IS applicators: (a) tandem and ovoid and (b) tandem and ring, both with IS option (Elekta, Stockholm, Sweden). (c) and (d) show the placement of these two applicators in patients, respectively.

to determine the most plausible uterine/cervix deformation resulting from a given intrauterine tandem. However, with large distortions caused by the applicators, such speculation to guide implant strategy is far from accurate and cannot systematically prevent suboptimal dosimetry post-insertion. Hence, this thesis aims to build predictive models to guide BT processes and decisions and improve BT's efficiency and quality in LACC.

*This thesis deals with the challenging problem of using deep-learning for predicting the large anatomical deformations from*

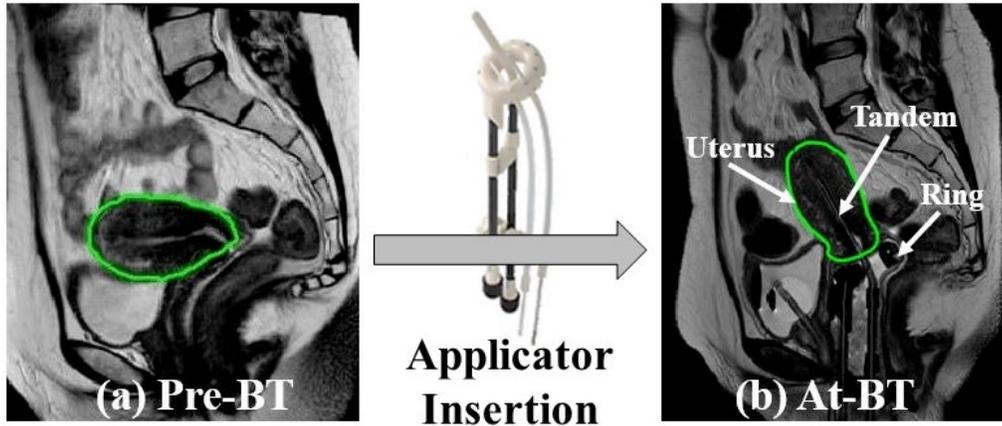


Figure 1.3: MRIs taken (a) prior to BT (pre-BT) applicator (intrauterine tandem and intravaginal ring; middle) insertion and (b) large/ complex uterine deformation post-applicator insertion (at-BT).

*pelvic MR images due to the insertion of a BT applicator to guide technical/clinical decisions in LACC BT.*

The large and complex deformations similar to Figure 1.3 cause problems for most intensity-based registration methods. In particular, the literature includes very few investigations of deformable image registration (DIR) procedures to handle uterus deformation for defining new dose accumulation models [13, 15, 16]. Deformations observed during EBRT and BT are often more unpredictable and challenging for DIR. Some intensity-based registration methods can register organs with such large deformations with the help of manually or automatically segmented MR images [17]. Studies have demonstrated that feature-based registration methods can also register large deformations in the abdominal and pelvic regions. However, none of the existing registration methods considered sliding between tissues [18, 19]. This complex problem can be simplified by aligning segmented tissues separately and generating a transformed dose distribution for each tissue independently. However, handling multiple-dose distributions for one treatment is not practical and not

supported by the treatment planning systems [20].

Machine learning (ML) techniques are used in various aspects of medicine, including radiation oncology. Ample evidence from recent publications explores its utility and future use in EBRT [21, 22]. ML has been applied for BT procedures ranging from decision-making to treatment completion. ML techniques have improved efficiency and accuracy by reducing human errors and saving time in certain aspects. Besides direct use in BT, ML contributes to contemporary advancements in radiology and associated sciences that can affect BT decisions and treatment. Existing imaging-related studies using Artificial Intelligence (AI) and Deep Learning (DL) can generally be divided into image enhancement, image registration, and image segmentation tasks in medical imaging. ML techniques may consolidate it further by reducing human effort and time. Prospective validation over more extensive studies and the incorporation of ML for a larger patient population would help improve the efficiency and acceptance of BT. The need for developers and researchers to rigorously validate ML-based tools and models, for users to understand how to operate them appropriately, and for organizations to develop guidelines for their use and maintenance is essential [23, 24].

Manual segmentation in regions without clear organ boundaries is complex, tedious, and time-consuming for clinicians; hence the need for automated delineation of the areas is of great interest. This study proposes a novel automated segmentation approach to delineate the uterus from MRI using a deep convolutional neural network (CNN) architecture. The proposed method incorporates autoencoder algorithms to detect the segmentation contours directly from the pre-BT MR images. Previous unsupervised approaches, such as principal component analysis using linear reconstruction [25], are unsuitable since they cannot preserve the highly nonlinear relationships when projected

to low-dimensional space. The proposed automated segmentation algorithm using autoencoders could be used for segmenting the closed region as well as open regions or open contours [26]. After segmenting the uterus automatically by CNN, the uterine deformation is predicted using a modified version of U-Net model. We proposed a novel approach to predict the deformed uterus structure using deep learning: a problem that has not yet been proposed in any of the studies before. This thesis also proposes a shape-based nonrigid registration method to handle complex and large deformations in the cervix-uterus structure for LACC patients treated with BT. The uterus, uterine canal, os, and vaginal canal were segmented manually on both the pre-BT MRI and at-BT MRI. As the BT applicator is inserted through the vagina into the uterus, both get deformed post-applicator insertion (i.e., at-BT). One of the main aims of this study is to predict this deformation. To achieve this, we must accurately register structures like the uterus, uterine canal, and vaginal canal on the pre-BT and at-BT images.

**This study was performed on paired pre-BT and at-BT MRI images from 120 LACC patients treated with EBRT, followed by a BT boost. The ultimate goal of this work is to develop a fast and consistent method that would allow for the prediction of the geometric changes experienced by the uterus in cervical cancer patients due to the BT applicator insertion.**

## 1.1 Motivation

The main aim of radiation therapy (RT) is to deliver a clinically effective dose to the target while minimizing the damage to the surrounding healthy tissues. Medical image processing is an integrated part of patient management

in modern RT. Image registration is used to find the spatial correspondence mapping between two image sets acquired at different times throughout the treatment [27, 28]. Accurate and reliable image registration techniques can improve the patient setup, estimate the dose delivered, and assess the possible anatomical deformations during the treatment. Fast image registration and organ deformation evaluation is required to improve the radiotherapy treatment for cervical cancer patients [29]. *However, registration of images in the pelvic region is very challenging due to the extreme organ displacements due to the insertion of the BT applicator and the large differences in bladder and rectum filling and positions during EBRT [30].* Several methods can handle the large deformations in the pelvic area [17, 18].

During cervical cancer treatment, target (tumor) volumes change position and shape due to organ motion, tumor regression, patient positioning, and applicator insertion in EBRT and BT. To address this challenging task, innovative deep learning (DL)-driven method was developed to predict uterine shape and location when deformed from its natural anatomy by the presence of a BT applicator. This technique uses only the pre-BT MRI - *a problem that has not been reported by any other study.*

Symmetric image registration methods were used to register intensity-manipulated images of prostate cancer patients, but their extension to images of cervical cancer patients is limited by the *complexity and the large magnitude of uterus deformation [31].*

Auto-contouring of different organs of interest aims to reduce delineation time and effort and to improve inter-observer consistency. Generating contours directly from deep learning techniques has emerged as a promising method of addressing automatic segmentation. DL-based contouring typically trains a convolutional neural network (CNN) directly from the data without any

need to identify the image features. Improved computing power and training of neural networks have made DL methods readily available for contouring purposes. Automated multi-organ segmentation is of great interest in medical image processing.

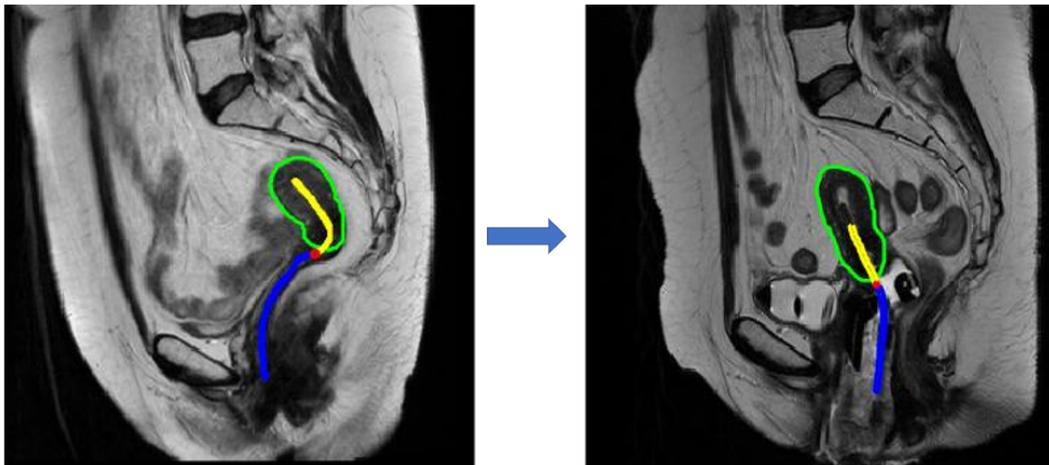
Insertion of BT applicators can dramatically alter the geometry of healthy tissues and target tumour structures (see Figure 1.3). Also, because of tumour regression during irradiation and variability in implant insertion and packing techniques, the bladder and rectum are likely to differ in shape and location relative to the applicators from insertion to insertion. This makes the problem even more challenging. The deformation of complex organs is still not entirely handled by DIR, which limits the ability to define accurate dose accumulation and toxicity models. In particular, the literature mentioned a few investigations into DIR methods to handle the deformation for defining new dose accumulation models [15, 16]. *There are no significant studies that quantified the deformation between BT images for cervical cancer. The primary goal of this thesis is to improve the BT treatment which is of great interest.*

This thesis aims to evaluate and optimize existing DIR methods to estimate the deformation in the most challenging scenarios seen in cervical cancer BT, precisely that of registering the patient anatomy between pre-BT and at-BT images. Modelling the deformations throughout treatment for LACC can significantly impact dose monitoring and planning optimization. We propose a deformable registration method using free-form deformation to map the uterus structure, the uterine canal, and the vaginal canal with and without an inserted BT applicator. *The quantification of uterus deformations during LACC treatment and their effect in terms of a local dose is currently an unsolved problem [32].*

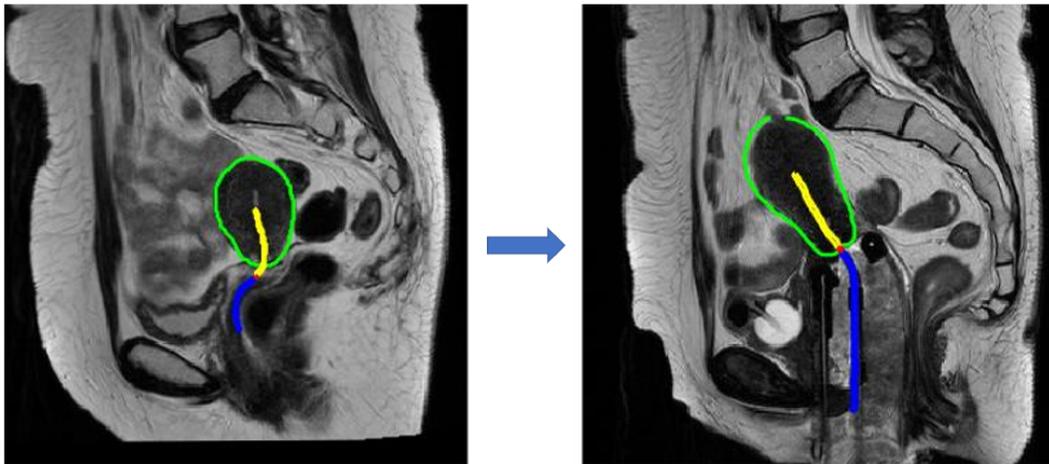
## 1.2 Challenges

The problem of predicting anatomical deformation due to applicator insertion is a very rare and difficult problem in the medical imaging field. The challenge in implementing our proposed algorithm was the inter-patient anatomical dissimilarity and extreme intra-patient uterine deformation from pre-BT to at-BT in the dataset. Increasing the size of our training dataset, including more heterogeneous images with anatomical variability, will improve the prediction accuracy of this DL-based algorithm. In most cases, the uterus moves or changes its position from left to right. A few of the most challenging and unusual cases or datasets are shown in Figure 1.4 and Figure 1.5, where the uterus changes uniquely for each patient. The main limitation of our dataset is that, we only have 2D MRI images of pelvic region (pre-BT and at-BT MRIs) to predict the uterine deformation due to applicator insertion in LACC BT. It is very challenging to develop a simulation model of the internal organs of pelvic region for one patient only from 2D MRI slice.

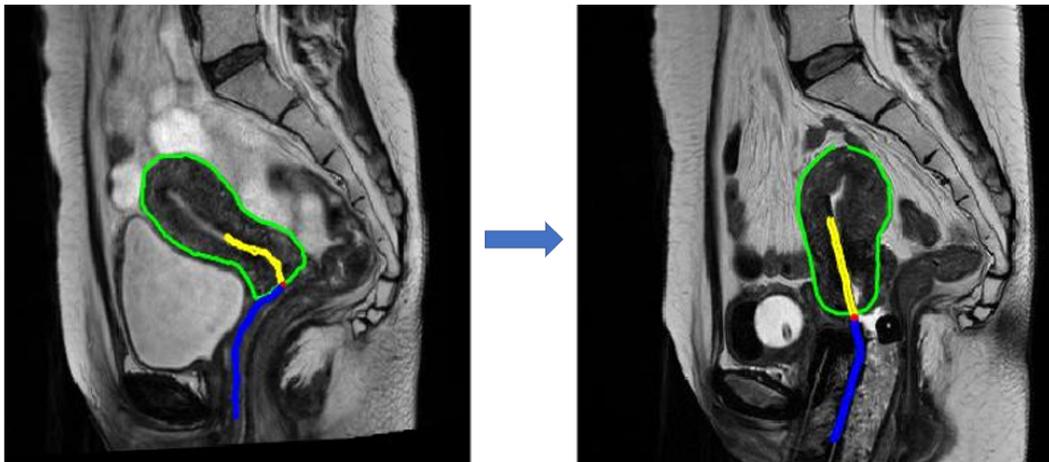
We propose to build predictive models to guide BT processes and decisions that can improve BT efficiency and quality in LACC. No previous study has quantified the deformation between pre- and post-applicator insertion images acquired for cervical cancer BT. Despite being trained on heterogeneous and complex deformations, our proposed DL-based model can successfully predict uterine distortion due to applicator insertion using only the pre-BT MRI. Furthermore, this approach can guide clinicians in selecting the most suitable applicator component and configuration ahead of the insertion procedure.



(a) Uterus shifts from extreme right to left

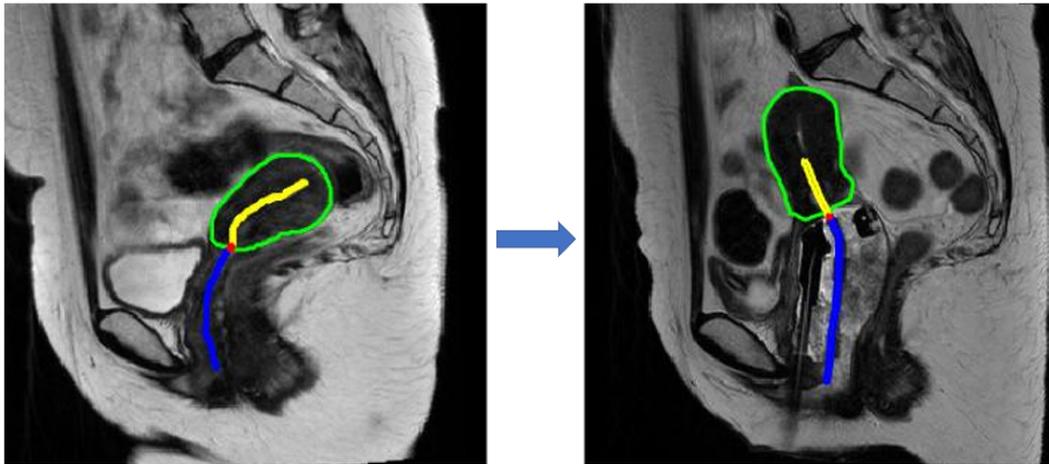


(b) Pre-BT uterus is elongated after application insertion

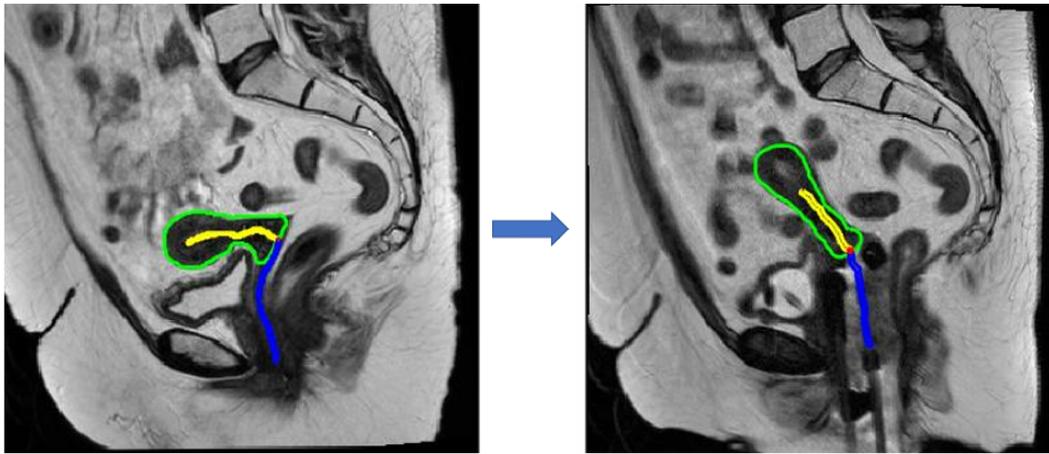


(c) Anatomy changes hugely from pre-BT state to at-BT state

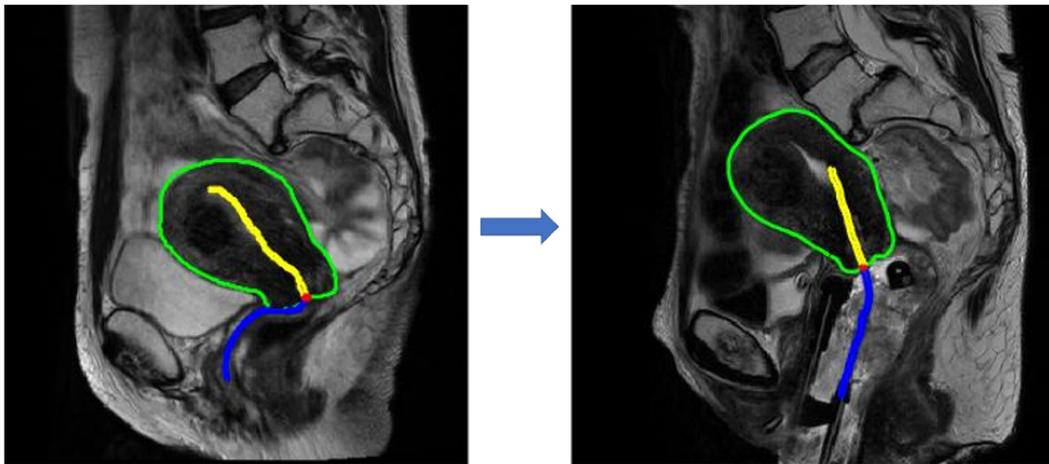
Figure 1.4: Some unusual and challenging datasets in our study where the uterus uniquely changes its position and shape



(a) Uterus is retroverted



(b) Uterus is comparatively thin and long



(c) Uterus is very big and bulky

Figure 1.5: Some unusual and challenging cases in our study where the size of the uterus varies greatly from patient to patient and showing the diversity in our dataset

## 1.3 Thesis Objectives

During cervical cancer treatment, target volumes change position and shape due to organ motion, tumour regression, patient positioning, and applicator insertion. To address this challenging task, we developed innovative DL-driven method to predict uterine shape and location when deformed from its natural anatomy by the presence of a BT applicator. This technique uses only pre-BT MRI — *a problem that any other study has not yet reported.*

The objectives of this thesis are:

- To perform a robust, intensity-based automatic image registration method to align the lumbar and pelvic spine (bone structure), the only rigid/fixed anatomy, in all the paired pre-BT and at-BT MRIs.
- To automatically delineate the uterus structure on MRI using a pre-trained deep convolutional neural network based on the Inception V4 architecture augmented with a multi-layer autoencoder.
- To automatically predict the large/complex deformations of the uterus using a modified version of U-net architecture. The predicted deformation is estimated using a free-form deformation algorithm.
- To demonstrate that DL can be used successfully to predict anatomical deformation to improve applicator component selection and configuration before the actual insertion procedure.
- To develop user-friendly software that clinicians can use for BT treatment planning and procedure.

## 1.4 Thesis Methodology

The overall framework to achieve our thesis goals contained several steps. First, compute optimal deformation parameters for the training data set as follows:

1. Perform a robust image registration to align the bone structures between all the pre-BT and at-BT MR-images in the dataset;
2. Train a deep-learning network to segment the uterus structure in all using all pre-BT Images automatically;
3. Predict the at-BT uterus shape and position using a convolutional neural network;
4. Measure or quantify the amount of uterine deformation due to BT applicator insertion;

### 1.4.1 Patient Information and Imaging Dataset

Pre- and at-BT MRI images of 120 cervical cancer patients treated between 2018-2019 were used to train all the parameters. All patients received EBRT followed by an IC or IC/IS BT boost using an interstitial ring and tandem applicator (Elekta, Stockholm, Sweden). The T2-weighted 2-D sagittal MRI were acquired on a 1.5T MRI scanner (Siemens Healthcare, Oakville, Ontario, (Headquarters: Erlangen, Germany).

### 1.4.2 Manual Segmentation/Ground Truth

Image pairs (pre- and at-BT) were imported into Oncentra (OcB), a BT treatment planning software (Elekta, Stockholm, Sweden). The uterus structure, uterine canal, vaginal canal, and external os were delineated manually by a

single Radiation Oncologist on the pre-BT images and a single Medical Physicist on the at-BT images. All contours were drawn on a single MRI slice on which the uterine and vaginal canals were visible. Figure 1.6 shows the pre- and at-BT MR image pairs along with the manually segmented contours of the regions of interest. The approximate time for delineation was 10 mins for the pre-BT and 5 mins for the at-BT images.

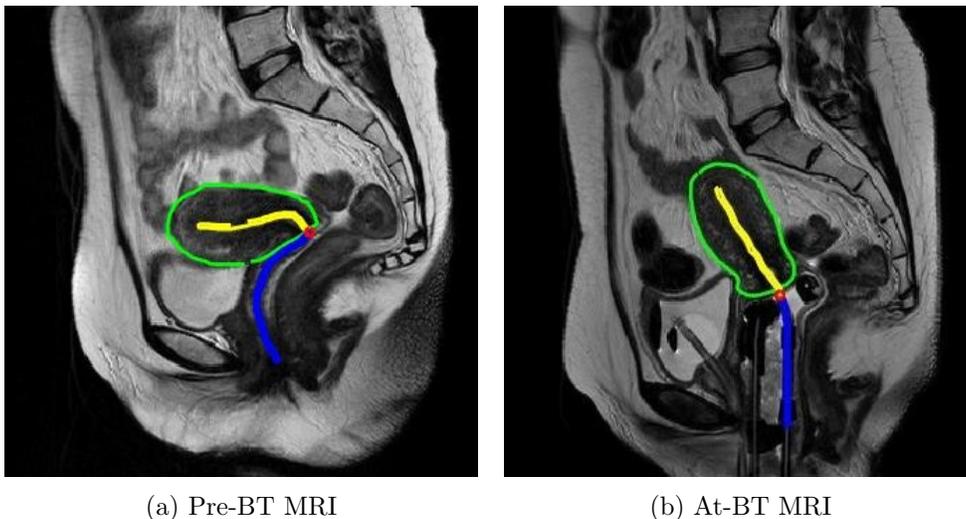


Figure 1.6: T2 weighted sagittal MR images of the same patient taken at (a) pre-BT and (b) at-BT. Both images show the 4 regions of interest used in this study. Green - uterus, yellow - uterine canal, blue - vaginal canal, and red - external os.

## 1.5 Clinical relevance

Predicting deformation from the natural anatomy before BT to anatomy in the presence of an intrauterine BT applicator is challenging as the uterus and surrounding organs deform uniquely relative to each other and for every patient. However, the applicator model and its geometry are fixed. Therefore, in clinical practice, radiation oncologists make several informed predictions before BT applicator selection and insertion. These predictions are based

on pre-BT MRI, gauging potential uterine deformation, the final selection of tandem relative to tumour targets, the size of the intravaginal applicator, and the radial and deep positions of the IS catheters. An implant strategy is then formulated for anatomic and geometric fit, which also considers anticipated dosimetry. The possibility of correctly predicting uterine distortion using deep learning methods is the first milestone towards achieving these technical BT parameters with improved accuracy and ease.

*Due to applicator insertion for personalised BT treatments, the proposed DL-based framework can be incorporated as an automatic prediction tool for uterine deformation. In addition, it promises more streamlined clinical/technical decision-making before BT applicator insertion resulting in improved dosimetric outcomes.*

## 1.6 Thesis Contributions

### 1.6.1 Contributions

- In this thesis, pelvic MRI data sets of pre-BT and at-BT from 120 patients were examined. The bony anatomy (lumbar and pelvic spine) was registered using the intensity-based automatic image registration method. Along with the bone structure, the position of the uterus was also aligned, which helps to measure the uterus deformation more accurately;
- A novel automated segmentation algorithm to delineate the uterus from MRI using a deep CNN augmented with autoencoder was developed successfully.
- Uterus deformation is predicted using a modified version of U-net ar-

chitecture. In this step, at-BT uterine deformation was predicted using only the pre-BT MRI image.

- Finally, the expected deformation is estimated or quantified using a non-rigid shape-based registration method with a free-form deformation algorithm.

## 1.6.2 Papers Published in Peered Reviewed Conferences

1. Shrimanti Ghosh, Kumaradevan Punithakumar, Fleur Huang, Geetha Menon, Pierre Boulanger, *Deep Learning Using Pre-Brachytherapy MRI to Automatically Predict Applicator Induced Complex Uterine Deformation*, IEEE Engineering in Medicine and Biology Society; July 2022; Glasgow, Scotland, UK.  
  
doi: 10.1109/EMBC48229.2022.9871157
2. Shrimanti Ghosh, Kumaradevan Punithakumar, Geetha Menon, Fleur Huang, Pierre Boulanger, *Predicting Brachytherapy Tandem-Related Uterine Deformation in Locally Advanced Cervical Cancer using Deep Learning with Free Form Deformation*, Women and Children's Health Research Institute; WCHRI Research Day November 2021; Canada.
3. Shrimanti Ghosh, Kumaradevan Punithakumar, Geetha Menon, Fleur Huang, Pierre Boulanger, *Deep Learning Predicts Uterine Deformation by Brachytherapy Tandem in Locally Advanced Cervical Cancer: A Step Towards Better Implant Strategies*, Canadian Organization of Medical Physics; June 2021; Canada.  
  
doi: 10.1002/mp.15102
4. Shrimanti Ghosh, Nilanjan Ray, Pierre Boulanger, Kumaradevan Punithakumar, Michelle Noga, *Automated Left Atrial Segmentation from Magnetic Resonance Image Sequences Using Deep Convolutional Neural Network with Autoencoder*, IEEE International Symposium on Biomedical Imaging; April 2020; Iowa, USA.  
  
doi: 10.1109/ISBI45749.2020.9098646

# Chapter 2

## Background

This chapter reviews various applications of artificial intelligence, machine learning, and deep learning in cervical cancer brachytherapy treatment. We will review the utilization of deformable image registration methods in brachytherapy and their limitations. Finally, we will address the problem of predicting uterine deformation in cervical cancer brachytherapy caused by the presence of the BT applicator. This particular application is almost non-existent in the literature, and no previous study has addressed this problem before - which makes this still an open problem in cervical cancer treatment and radiation oncology.

### **2.1 Application of Artificial Intelligence in Brachytherapy**

The advancement of artificial intelligence (AI) in medicine and, more specifically, in modern radiation oncology (RO) will impact all aspects of RO. More specifically, Machine learning (ML) and Deep learning (DL) are currently being incorporated into various aspects of medicine, including oncology. For

example, these techniques have applications in developing more efficient external beam radiotherapy machines. However, the discussion on its role in brachytherapy is sparse in today's literature. The main goal of this literature review is to summarize the available applications and discuss the potential uses of ML/DL in BT, including future directions. ML has been applied for BT procedures from decision-making to treatment completion. ML has improved efficiency and accuracy by reducing human errors and saving time. Besides direct use in BT, ML contributes to contemporary advancements in radiology and associated sciences that can affect BT decisions and treatment. Incorporating ML technologies for a larger patient population would help improve the efficiency and acceptance of BT.

### **2.1.1 AI in Brachytherapy Pre-Planning**

Pre-planning is an essential component in BT that includes review of the clinical situation, assessment of volume to be treated, approach and technique to be used for implanting the tumor, choice of applicator and planning a prescription, depending on the surrounding vital structures. AI can play a significant role in this case. We know that DL can perform fully automatic segmentation from MRI images with good accuracy and in a clinically feasible time [33]. AI has been utilized in low dose rate (LDR) seed BT of prostate cancer. In the literature [34], authors used ML to extract dosimetrically optimal pre-planning methods which were comparable in quality to those by expert planners, but with a significant reduction in planning time ( $0.84 \pm 0.57$  minutes vs  $17.88 \pm 8.76$  minutes,  $p = 0.020$ ). There are preliminary reports that DL methods using previous experiences can guide selection of suitable applicators for high dose rate (HDR) BT in cancer cervix [34]. This has been validated in the choice of interstitial over intracavitary applicators based on geometric characteristics of

data such as shape and volume of high-risk clinical target volume. ML/DL algorithms can help in decision-making and augmenting a physician’s judgment leading to more consistency, obviating many logistic issues and last-minute unwanted plan changes in the operating room, with no compromise in plan quality [35].

Accurate prediction of clinical outcomes of cervical cancer could help adjust or optimize the treatment of cervical cancer and benefit the patients. Statistical models, various medical images, and machine learning have been used for outcome prediction with promising results. Compared to conventional statistical models, machine learning has demonstrated advantages in dealing with the complexity of large-scale data and discovering prognostic factors. It has great potential in clinical application and in improving cervical cancer management. However, the limitations of prediction studies and models include simplification, insufficient data, overfitting, and the lack of interpretability. In many applications, more work is needed to make clinical outcome prediction more accurate, reliable, and practical for clinical use. More specifically, deep learning (DL) networks have been applied extensively in many fields due to the introduction of fast and inexpensive computer hardware (GPU) and the availability of large annotated databases. As a result, DL has gradually emerged in BT treatment [36], focussing on organ delineation, applicator reconstruction, dose calculation and treatment planning systems [37, 38].

### **2.1.2 AI in assisting procedure in BT operating room**

Diagnostic ultrasonography (USG) is a powerful tool in real-time BT for guiding interstitial needle insertion. (3D) USG image analysis has shown great potential in USG-based clinical application of BT [39]. Application of novel DL in automated imaging analysis tasks (lesion classification, organ segmenta-

tion, object detection, registration, measurements, quality assessment) help in treatment planning procedure in BT operating room. AI application of USG in future may provide real-time guidance in operating room to identify target as well as critical structures in patients under going implants for BT [39]. Besides USG, endoscopy is an under explored modality for seed or catheter placement. There are case reports of using AI-driven navigation system for real-time localization of the airways and lung nodules using fluoroscopic images [40]. LDR seeds may be placed via endoscopic routes in several sites such as pancreas, lungs; usage of AI here may help in more accurate placement of radioactive applicators with more efficacy. Whether these scenarios translate into real- world applications of AI in guiding BT procedures will become clear only with wider availability and time. At present, the limited experience with USG and endoscopy in the hands of RO professionals indicates that this AI approach, although promising, would be slow to be adopted [40].

## **2.2 Organ Delineation Using Deep Learning**

In BT, organ delineation and segmentation undoubtedly play an essential role in the treatment plan. In clinical practice, manual segmentation is still the primary method. However, automatic segmentation from medical images plays an increasingly important role in helping clinicians delineate tumour areas and organs at risk (or normal tissue). Many segmentation methods have been proposed in the literature, which differ according to their application and imaging modality.

### 2.2.1 Automatic Segmentation in Cervical Cancer using Deep Learning

Brachytherapy (BT) is an essential part of the curative intent therapy and closely associated with improvements in clinical outcomes. Three-dimensional (3D) image-based BT allows individual treatment planning based on the volumetric image of patient and is considered as a significant technical advancement and widely adopted for the treatment of cervical cancer. The application of 3D image-based BT enables the practitioner to prescribe dose to the target volume as well as determine and potentially limit dose to the organs at risk (OARs), which is more advantageous than the conventional two-dimensional (2D) image-based approach. Numerous studies demonstrate improved treatment plan quality and clinical outcomes of 3D image-based BT for cervical cancer. Magnetic resonance imaging (MRI) is the preferred imaging modality for treatment planning of cervical cancer BT due to its superior soft tissue visualization relative to computed tomography (CT). However, there are many obstacles for routinely performing the MRI-based BT in many radiation oncology departments, including limited availability, high cost, and long scanning time [41]. Therefore, CT-based BT of cervical cancer is widely used in treatment centers worldwide, especially in the developing countries.

In [41], the authors proposed a novel 3D CNN architecture that is based on the popular 3D U-Net architecture<sup>30</sup> with incorporation of residual connection, dilated convolution and deep supervision (henceforth referred to as DSD-UNET). The proposed network is trained and evaluated for automatic segmentation of high-risk clinical target volume (HR-CTV) and OARs in the planning CT of cervical cancer BT. Performance of the DSD-UNET is then compared with that of the conventional 3D U-Net. One of the limitations of this work is the relatively small dataset size. This is due to the limited number

of patients with cervical cancer that received CT-based BT in our clinic and the lack of common dataset that is suitable for this segmentation task. To ease this problem, data augmentation strategy was applied in the model training. Dropout was deployed in the network to reduce the risk of overfitting introduced by data augmentation. However, due to the intrinsic characteristics of deep learning method, larger dataset usually leads to the improvements of performance and generalization. Therefore, we plan to collect more suitable image data in the future study. Then more accurate and reliable segmentation result could be expected. The most inferior segmentation performance for the DSD-UNET model was observed on the segmentation of sigmoid, with the lowest mean DSC value of 64.5% and higher mean HD value of 19.6 mm among all structures.

Accurate prostate segmentation is key to biopsy needle placement, BT treatment planning, and motion management. Lei [42] et al. used CNN to develop an automated, accurate, and stable segmentation method to delineate the clinical target volume using transrectal ultrasound images of the prostate during BT. However, manual segmentation during these interventions is time-consuming and subject to inter- and intra-observer variation. To address these drawbacks, the authors proposed a DL-based multidirectional approach which integrates deep supervision into a 3D patch-based V-Net for prostate segmentation. This 3D supervision mechanism deals with optimization difficulties when training a deep network with limited training data. During the segmentation stage, the patches are extracted from the newly acquired ultrasound image as the input of the well-trained network, and the well-trained network adaptively labels the prostate tissue. The final segmented prostate volume is reconstructed using patch fusion and refined through contour refinement processing. The Dice similarity coefficient (DSC) and Hausdorff distance (HD)

were  $0.92 \pm 0.03$  and  $3.94 \pm 1.55$ , respectively.

Qin Nannan [43] et al. completed the automatic delineation of clinical target areas and organs at risk in BT of cervical cancer by building a U-net network. The average value of the Dice similarity coefficient of the automatically delineated target area was 0.89, and the average Hausdorff distance was within 5.3 mm, which proves that it can be used in the clinic and can significantly improve contouring efficiency.

## 2.3 Applicator Segmentation and Reconstruction

Applicator design is critical during BT treatment as they assist in placing the radioactive source at the desired location to deliver the prescription dose at the tumour locations. Hence, it is imperative to accurately design these applicators as part of the treatment plan to deal with source distribution and dose optimization. Currently, clinicians segment the BT needles and applicators manually during the treatment, which is time-consuming and error-prone. So, a modified deep U-Net is used to segment the pixels belonging to the BT needles. In addition, VGG-16-based deep CNN is combined with the segmentation network to predict the locations of the needle tips [44]. This helps to improve the efficiency and quality of BT treatments.

Applicator reconstruction is the process of localizing the radiation source paths defined by the applicator channels in the planning images. It is another critical step during the procedure of BT treatment planning. The potential dwell positions are placed on the digitized applicator channels and corresponding dwell times are determined to meet the dosimetric objectives. Applicator reconstruction accuracy has a significant impact on the dosimetric result of

the treatment plan due to the steep dose gradients of BT treatment. A small uncertainty in the digitization of applicator channels would translate into a relatively large dosimetric uncertainty. In general, applicator reconstruction is performed manually by the medical physicist. The digitization process is subjective and time-consuming. Thus there is a strong need to achieve fully automatic applicator reconstruction in 3D image-based BT to ensure treatment planning accuracy and efficiency. The applicator library integrated in the treatment planning system is the clinically available tool for automatic applicator digitization, which can significantly reduce the reconstruction uncertainty and improve efficiency. It allows channel digitization based on the manual registration of virtual applicator model with predefined source paths to its appearance in the planning images [45]. However, the applicator library-based reconstruction method is not fully automatic due to the manual alignment of applicator model. Moreover, applications of this method are limited to only those applicators included in the library. Electromagnetic tracking technique has been recently utilized for catheter digitization in BT. Although this method has highly accurate digitization result, additional hardware and complex procedure may hamper its widespread application.

Segmentation results for all parts of the tandem and ovoid applicator using the DSD-UNET model were assessed. It is observed that outstanding segmentation accuracies were achieved for all parts of the applicator. The mean DSC and JI (Jaccard Index) values for all the applicator components were higher than 88.0% and 80.0%, respectively. In particular, automatic segmentations of the intrauterine tube and ovoid tubes achieved superior performances compared with those of the other applicator components (average DSC value of 92.1%, average JI value of 86.8%, average HD value of 2.3 mm). The best segmentation accuracy was observed on the segmentation of intrauterine tube

(DSC = 92.6%, JI = 87.7% and HD = 1.9 mm). Quantitative evaluation results show that the proposed DSD-UNET method outperformed the 3D U-Net and could segment the HR-CTV, bladder, and rectum with relatively good accuracy.

## 2.4 Dealing with Large Deformations in Cervical Cancer BT Imaging

The first step toward estimating the total EBRT + BT dose distribution is correctly aligning the underlying anatomy. This task is particularly challenging for cervical cancer treatments due to the large and complex deformations caused by tumour shrinkage, bladder filling changes, and insertion of a BT applicator [29]. Anatomical deformations can be predicted to a large extent by statistical or biomechanical deformable models [46, 47]. Image registration can be categorized into two groups: rigid and non-rigid. Non-rigid image registration is also known as deformable image registration (DIR). In rigid image registration (RIR), all pixels can move or rotate uniformly so that every pixel-to-pixel relationship remains the same before and after transformation.

Although there is no boundary to implementing DIR in the radiation therapy (RT) field, the application of DIR could be categorized into four major areas: dose accumulation, mathematical modelling, automatic segmentation, and functional imaging. The result of deformable image registration is a deformation vector field which describes the correspondence between the fixed and moving images. The known correspondence information between the two image sets is essential for validating DIR algorithms. DIR mainly deals with tasks such as complex non-linear and local distortion management, multi-modality image registration and multi-dimensional image registration, all of which make

automatic image registration particularly challenging.

Manual delineation of clinical target volumes remains a time-consuming task in radiation oncology. However, some research works propose auto-delineating tumour volume using atlas-based registration techniques [48, 49]. In [50], distance metrics can provide sufficient information to automate the delineation of high-risk target volumes. Using deep auto-encoders [51] provide a venue for good generalization even when a few patient images are used for training. In [52], a deep learning approach in which the model is trained on anatomic structure distance map information has been shown to produce patient-specific high-risk clinical target volumes.

DIR methods are mainly used to register the images between EBRT and BT. But no such method is described in the earlier literature for measuring or predicting the anatomy deformation during the BT treatment in cervical cancer. Therefore, we are trying to solve this unique problem in radiation therapy.

## 2.5 Various Non-rigid Image Registration

Image registration aims at finding the optimal transform that best aligns structures in two input (2D) or volume data (3D) images. It is an important part of image analysis and used in several disciplines. In the medical field, image registration is a key component in several areas including the fusion of morphologic and functional images, intervention planning, computeraided diagnosis (CAD) and treatment follow-up [53], atlas building [54], radiation therapy [55], model-based segmentation [56], and computational model building [57]. The choice of deformation model is of great importance for the registration process as it entails an important compromise between computational efficiency and

richness of description. It also reflects the class of transformations that are desirable or acceptable, and therefore limits the solution to a large extent. The parameters that registration estimates through the optimization strategy correspond to the degrees of freedom of the deformation model. Increasing the dimensionality of the state space results in enriching the descriptive power of the model. This model enrichment may be accompanied by an increase in the model's complexity which, in turns, results in a more challenging and computationally demanding inference. Furthermore, the choice of the deformation model implies an assumption regarding the nature of the deformation to be recovered.

In general, an important drawback of registration is that when source and target volumes are interchanged, the obtained transformation is not the inverse of the previous solution. In order to tackle this shortcoming, Christensen and Johnson [58] proposed to simultaneously estimate both forward and backward transformations, while penalizing inconsistent transformations by adding a constraint to the objective function. Linear elasticity was used as regularization constraint and 3D Fourier series were used to parameterize the transformation.

Stefanescu et al. presented a way to perform adaptive smoothing by taking into account knowledge regarding the elasticity of tissues in [59]. A non-stationary diffusion filter was used to smooth less inside areas where greater deformations were expected and smooth more inside objects where coherence should be preserved. The authors also proposed to take into account the local image gradient content during smoothing. In areas with large image gradients where the local confidence for the established correspondences is higher, smoothing is scaled down. On the contrary, smoothing is scaled up in homogeneous areas.

## 2.6 Diffeomorphic Image Registration

Image registration is an essential task used in many medical image analysis applications such as assessing disease progression over time, merging and comparing different image modalities, and shape analysis. By maximizing the image similarity, such as intensity correlation, image registration provides the correspondence and non-linear transformation between pairs of images. Diffeomorphic image registration offers more desirable properties such as smooth deformation, topology preservation, and transformation invertibility. Diffeomorphic mapping is the underlying technology for mapping and analyzing information measured in human anatomical coordinate systems which have been measured via Medical imaging [60]. Diffeomorphic mapping is a broad term that actually refers to a number of different algorithms, processes, and methods. It is attached to many operations and has many applications for analysis and visualization. It can be used to relate various sources of information which are indexed as a function of spatial position as the key index variable. Large deformation diffeomorphic metric mapping (LDDMM) is a specific suite of algorithms used for diffeomorphic mapping and manipulating dense imagery based on diffeomorphic metric mapping within the academic discipline of computational anatomy, to be distinguished from its precursor based on diffeomorphic mapping [61]. LDDMM delivers a non-linear smooth transformation with a favorable topology-preserving one-to-one mapping property [61].

Successful segmentation algorithms built on automatic to semi-automatic algorithms making use of some form of expert prior knowledge are likely to overcome the mentioned difficulties. The LDDMM algorithm provides non-rigid registration between two grayscale anatomical images, which can be used to develop automated segmentation methods. Given a template image, target im-

age, and template segmentation, the non-rigid mapping of the template image to the target image can be used to propagate the template segmentation, generating a target segmentation. In [62], authors proposed semi-automated ganglia segmentation method using LDDMM. The semi-automated segmentation scheme is minimal and does not require expertise in brain anatomy. The user is necessary for four tasks: (1) automated thresholding-based segmentation of the ventricles, (2) landmarking of the ventricular surfaces, (3) definition of the bounding box to delineate the image region for matching, and (4) identification of average caudate intensities for intensity equalization. The automated and the manual segmentations are found to differ on the exterior boundary of the caudate. Due to the elongated, narrow shape of the caudate nucleus, the ratio of the number of voxels on the surface of the caudate to the total number of caudate voxels is very high, on average being 71%. Therefore, the partial volume effects are likely to be a heavy influence in calculation of the L1 distance which measures the difference in voxel labeling between the automated and the manual segmentation, especially since there are a large number of these on the exterior of the caudate that are not exactly matched due to regularization constraints placed in computing dense diffeomorphic transformations.

Another multistructure framework has been introduced in [63], with concurrent subcortical and cortical shape matching to guide the overall image registration using diffeomorphic image registration. The significant advantage of being flexible enough to allow use of computed features derived from MRI images, such as white matter tract-based labels and diffusion tensor metrics or additional modalities such as BOLD functional images, susceptibility weighted images, magnetization transfer images, or quantitative relaxation-based maps. A multicontrast LDDMM approach specific to the problem of diffusion tensor image registration has also been proposed. At first, the im-

ages are aligned using affine registration and intensity normalization on the Freesurfer tissue segmentation images of the template and target images. To limit computational and memory requirements, a bounding box is constructed around the template brain. After that diffeomorphic image registration was applied. Tensor based Morphometry was applied to investigate how well the multistucture and MRI-only registration methods can detect the expected differences between a group of demented patients and cognitively normal patients. This approach of using multiple structure segmentations to drive local whole brain registration shows for the first time that, as expected, incorporation of individual segmentations into the registration further improves local registration accuracy. This work showed that a group-wise average atlas built with multistucture registration accounts for greater intersubject variability and provides more sensitive tensor-based morphometry measurements. One limitation of this method is large amount of computational resources required for multistucture registration.

## 2.7 Modeling Soft Tissue Deformations

Large tissue deformation frequently occurs during many medical interventions. An accurate simulation of these procedures necessitates considering these tissue displacements by modelling tissue deformation as it interacts with medical tools. Deformation modelling has been an active topic in computer graphics for fitting noisy data and the simulation of clothing, facial expressions, human/animal characters, surgical simulators, etc. Although most of those methods are based on mass-spring models because of their simplicity and computational requirements, others use the Finite Element Method (FEM) to deal with the complexity of soft tissue material properties. The main limitation of

applying these methods to our problem is that we do not have the 3D images and the corresponding ground truth segmentation. The main reason we concentrated on a 2D dataset is that clinicians currently rely on the 2D MRI data for the BT treatment. Current research on deformable models focuses mainly on the object’s surface [64] using geometric or non-physical models for computer animations. Because of these limitations, they are not accurate enough for medical applications. Other approaches using physics-based models can approximate the non-linear behaviour of soft tissue more accurately. However, their high computational and storage requirements are problematic for clinical applications. These models include mass-spring models (MSM) [65,66], boundary element method (BEM) [64], meshless method [67] and the widely used Finite Element Method (FEM) [68]. Most of these methods consider soft-tissue deformations based on the theory of elasticity and employ constitutive equations to account for the complex mechanical behaviour of soft tissues.

Plantefeve proposes a co-rotational FEM [69] algorithm based on linear co-rotational elasticity to accommodate the non-linear deformation of soft tissues. However, it can only handle geometric rather than material nonlinearity as it uses a linear material deformation law. Another algorithm proposed by Miller [70] uses the total Lagrangian explicit dynamics FEM to account for both geometric and material nonlinearities involved in soft tissue deformation. However, this method does not allow changes in model topology in the simulation since it applies pre-computation for spatial derivatives. The main disadvantages of FEM algorithms are that they require a large amount of data to represent the mesh used in nodal connectivity and other parameters. Therefore, they usually need longer computational time, the need to initialize physical property parameters etc.

A new neural network methodology for modelling soft tissue deformation

for surgical simulation by Jinao *et al.* [71]. It formulates soft tissue deformation and its dynamics as the neural propagation and dynamics of cellular neural networks for real-time, realistic, and stable simulation of soft tissue deformation. Two CNN models are developed based on 3-D volumetric tissue models to carry out soft tissue deformation via neural propagation and dynamics by combining bioelectric energy propagation of soft tissues with non-rigid mechanics of motion. One CNN model is developed by formulating the cell current source as the strain energy density and the local connectivity of cells as the Laplace operator to describe the process of soft tissue deformation as the bioelectric propagation of mechanical load in soft tissues. The other CNN model is developed based on non-rigid mechanics of motion to govern the model dynamics of soft tissue deformation. The proposed methodology not only satisfies the real-time computational requirement of surgical simulation due to the collective and simultaneous activities of cells but also achieves the physical behaviour of soft tissue deformation due to the neural propagation of mechanical load. Furthermore, it performs stable model dynamics for soft tissue simulation but with similar computational efficiency to explicit time integration. The integration with a haptic device is achieved for interactive soft tissue deformation with haptic feedback. Both methods and experiments tested a volumetric kidney model on a lamb kidney to measure the force-displacement data. A digital force gauge was used to apply a compressive force to the lamb kidney, and its displacement was controlled and recorded by a micrometre barrel. The kidney specimen was placed in a plastic tray with oil applied to the interface of the plastic tray and tissue to minimize friction. The trials were also conducted using haptic feedback on a virtual human kidney model. The computation time was increased with the increase in the number of points. A stable and reliable deformable model to simulate the dynamics of

soft tissues is a challenging area of research. In this paper, Duan [72] propose a soft tissue simulator of the liver and gallbladder to deform the organs and interact with virtual surgical instruments accurately based on a mass-spring model (MSM). This paper presented a soft tissue simulator that uses a fast tetrahedral mass-spring formulation to calculate soft tissue deformation based on explicit integration. The model parameters are selected using soft tissue properties. Again, to model solid 3D objects like the liver and biliary system, a tetrahedral mesh modelling with both the surface and internal structure is implemented. Firstly, they simulated the porcine liver deformation with the obtained MSM and compared it with the non-linear FEM. Different mesh models and parameters were tested, and the deformation of the MSM was always close to the FEM reference. Secondly, the MSM was applied to the porcine liver and gallbladder system. The simulation was directly compared with the actual tissue deformation, and good approximations were also obtained for the model consisting of multiple different materials. Finally, a real-time simulation was developed based on the obtained MSM. They did not consider any friction force between the instrument and soft tissues and the GPU techniques to speed up the MSM-based physical engine in this work.

Most of these physical-based simulation methods for soft tissue deformation models 3D data or any newly harvested porcine liver/gallbladder are a little challenging to use and beyond the scope of our project. For that reason, we concentrated our effort on 2D data and the problem of simulating the deformation of the uterus.

### 2.7.1 Physical methods to simulate deformable materials

In the last decades, the computer graphics community has developed a wide variety of physically based models to address the challenge of simulating natural elements and deformable materials. For the latter, constitutive laws are used for the computation of the symmetric internal stress tensor, and a conservation law gives the final partial differential equation (PDE) that governs the dynamics of the material [73]. A majority of simulation methods in computer graphics use 2D and 3D meshes [74]. Most of these approaches are based on mass-spring systems, or the more mathematically motivated Finite Element Modeling (FEM), Finite Difference (FD) or Finite Volume (FVM) Methods, in conjunction with elasticity theory [74]. In mesh based approaches, complex physical effects, such as melting, solidifying, splitting or fusion, pose great challenges in terms of restructuring. Additionally, under large deformations the original meshes may become arbitrarily ill-conditioned. For the simulation of these complex physical phenomena, efficient and consistent surface and volume representations are needed, which allow simple restructuring. Most of the physical models are derived from continuum mechanics, which allows the specification of common material properties such as Young's Modulus and Poisson's Ratio [74].

However, continuous systems have infinite degrees of freedom, making their description difficult for geometric and dynamic aspects. In mathematical terms, we are dealing with infinite basis functions, maybe uncountable. One possibility to simplify the problem is considering finite-dimensional representation with enough flexibility to represent the solution with the desired precision. In the context of mechanical systems, the finite element method (FEM) is the traditional way to perform this task. When simulating a deformable body, the

3D/2D object's geometry is usually represented by mesh-based methods that support FEM-based techniques [75].

### 2.7.2 Mass Spring Model (MSM)

Another possibility for elastic object simulation is to apply discrete models based on mass-spring models (MSM). In this case, the object's geometry is represented by a mesh, and its nodes are treated like mass points while each edge acts like a spring connecting two adjacent nodes. MSM are simple to implement and can be faster than continuous ones, so, more suitable for real-time applications. Therefore, MSM techniques have been used to model deformable objects for woven cloth simulation, and soft organic tissues, like muscles, faces or abdomen in virtual surgery applications [76–78].

However, the main limitation of the MSM is the difficulty of designing them to represent the mechanical behaviour of deformable bodies with enough accuracy [79]. The relation between mass-spring models and the continuum elasticity theory was examined in references [79, 80]. The conclusion is that methods that are based on continuum mechanics are, in general, more realistic than their discrete counterparts. This happens because mechanical systems depend on their macroscopic parameters (Young's module and Poisson's ratio) and constitutive equations that characterize the nature of the materials that make up the bodies. However, there is no general physically based or systematic method in the literature to determine the mesh topology or MSM parameters from known constitutive behaviour [80].

Mass-spring models (MSM) are frequently used to model deformable objects for computer graphics applications due to their simplicity and computational efficiency. However, the model parameters are not related to the constitutive laws of elastic material in an obvious way. The MSM parameters

computation from a model based on continuum mechanics is a possibility to address this problem.

### **2.7.3 Application of MSM in Brachytherapy**

A stable and accurate deformable model to simulate the dynamics of soft tissues is a challenging area of research. Real-time and precise simulation of soft tissue deformation is still a significant challenge. A realistic estimation of stiffness parameters (Young's modulus, shear modulus, bulk modulus, viscosity) is required for the simulation models. The reliable approximation of these input parameters is difficult. FEMs and MSMs are most suitable when we have the appropriate data and the accurate values of the parameters responsible for the anatomy deformation.

As our original problem predicts uterine deformation, various factors affect soft tissue deformation. The elasticity property of the uterus, surrounding fluid density, and the presence of other organs affect the deformation massively from patient to patient. We have seen from our data that every patient's anatomy is different and deforms uniquely. So, predicting this deformation before applicator insertion is challenging even for expert clinicians. For that reason, we wanted to utilize deep learning methods and see if DL networks can discover the hidden information from the MRI data, with which we can come up with a method to predict the anatomy deformation in cervical cancer brachytherapy.

## 2.8 Limitations of Various DIR Algorithms in Brachytherapy

Large organ displacements, deformations, and volume variations occur between the pre-BT and at-BT anatomies. The quantification of the total delivered dose during BT needs to be investigated for LACC to understand better and predict OAR (organs at risk) toxicity risks [81]. A limited number of DIR methods have been proposed for registering pre-BT and at-BT images [14]. In [20], the authors proposed a workflow to map pre- and during BT anatomy where each organ was independently aligned using the symmetric thin plate spline robust point matching (sTPS-RPM) proposed in [18,19]. The studies reported few details about the impact of the BT applicator insertion and how to predict the deformation created by the applicator insertion. The biomechanical model-based DIR algorithm MORFEUS proposed by Brock [82] relies on contours only. It allows the generation of a dense deformation vector field (DVF) from the reference image to the moving image coordinates. Triangular surface meshes of the organ delineations are generated, and a deformable mesh registration method is used to determine boundary conditions. Using FEM, all structures of interest are built to solve the displacements of all the volumetric mesh nodes to generate a dense DVF. This method was applied to an MRI of the thorax and abdominal regions at a normal inhalation and exhalation breath hold. The amount of anatomy deformation that occurred is much less than the one encountered in BT deformation. More complex biomechanical models still need to be investigated to simulate more complex behaviour such as sliding, and applicator insertion/removal [14]. One of the unknowns in many of these algorithms is the biomechanical model's parameters that go beyond the simple linear elastic model used in most FEM algorithms. *To date, no*

*accurate biomechanical model-based approach has been proposed to predict the anatomy deformation due to applicator insertion.* The main limitation of some of the intensity-based DIR methods is they cannot handle complex deformation as it relies on the intensity information alone, like the well-known Demons algorithm [83]. A deformable female pelvis phantom (uterus and bladder) was created based on a patient's CT data using 3D printing [84] technology. Then the applicator was inserted into the phantom to simulate BT. After that, a hybrid intensity and structure-based DIR was used. The main limitation of all DIR or non-rigid registration methods is that they require both pre-BT (moving image) and at-BT (fixed image) images to deform the anatomy. The ability of the DIR method to separately quantify the deformation of the cervix and uterus is of high interest in radiation treatment [14], and no research has addressed the challenging task of predicting the deformed anatomy due to the applicator insertion [14]. Several publications have reported DIR techniques to fuse multimodality images to assess dosimetric changes. The novelty of this study is that it addresses the challenging task of predicting the deformed anatomy following applicator insertion from the natural anatomy before BT. When applied to the clinic, this unique method will be able to predict the deformed anatomy employing just the pre-BT image, a problem that has not been addressed yet by any other study.

# Chapter 3

## Automated Intensity-based Image Registration

### 3.1 Overview

Image registration, also known as image fusion, can be defined as the process of aligning two or more images. An image registration method aims to find the optimal transformation that best aligns the structures of interest in the input images. In our application, image registration aims to automatically establish correspondences between different images displaying views of objects or organs. These images may be acquired at different times, from different devices or perspectives, or even reveal different types of information. Many applications require nonlinear (i.e., not necessarily linear) alignment strategies; hence, nonlinear registration enters into play. There is a large number of application areas requiring image registration, and image registration has an impact on basically every imaging technique. Furthermore, the growing variety of medical imaging datasets has progressed with the parallel development of many image registration techniques. Therefore, image registration techniques

may be classified in various manners.

Medical image registration algorithms can be divided into three steps:

- Determine transformation between the source image and target image;
- Measure the similarity degree of the source image and target image;
- Adopt some optimization methods, making the similarity measure degree better and faster to reach the optimal value.

Image registration aims to apply a transformation operator  $g$  between two images  $I_1$  and  $I_2$  that optimized a similarity function  $E$  in order to find the optimal operator  $g_{opt}$ :

$$g_{opt} : I_1 \rightarrow I_2 = \min(E(I_1-g(I_2))) \quad (3.1)$$

## 3.2 Image Registration Methodology

Image registration applications can be divided into four main groups according to the manner the image was acquired:

- *Different viewpoints (multi-view analysis)*: Images of the same scene are acquired from different viewpoints. The aim is to gain a larger 2D view or a 3D representation of the scanned scene.
- *Different times (multi-temporal analysis)*: Images of the same scene are acquired at different times, often regularly, and possibly under different conditions. The aim is to find and evaluate changes in the scene between consecutive image acquisitions.
- *Different sensors (multi-modal analysis)*: Images of the same scene are acquired by different sensors. The aim is to integrate the information

from different source streams to gain a more complex and detailed scene representation.

- *Scene to model registration*: Images of a scene and a model of the scene are registered. The model can be a computer representation of the scene, for instance, maps or digital elevation models (DEM), another scene with similar content (another patient) etc. The aim is to localize the acquired image in the scene/model and/or to compare them.

There are four steps that every image registration algorithm must do. They are:

- *Feature detection*: A domain expert detects salient and distinctive objects (closed boundary areas, edges, contours, line intersections, corners, etc.) in both the reference and sensed images.
- *Feature matching*: It establishes the correlation between the features in the reference and sensed images. The matching approach is based on the content of the picture or the symbolic description of the control point set.
- *Estimating the transform model*: The parameters and mapping functions are calculated, which align the detected picture with the reference image.
- *Image resampling and transformation*: The sensed image is transformed using a mapping function. The appropriate interpolation technique computes image values in non-integer coordinates.

Image registration methods can be classified into area-based and feature-based methods. Therefore, various image registration methods are briefly discussed here.

- **Rigid registration:** In rigid image registration, the objects that need to be aligned can be translated and rotated with respect to one another to achieve correspondence. In rigid image registration, the rotated images are superimposed, and registration is achieved.
- **Nonrigid registration:** Nonrigid image registration refers to a class of algorithms where the images to be registered have geometric differences that cannot be accounted for by similarity (global translation, rotation, and scaling) transformations. These methods are capable of aligning images where correspondence cannot be achieved without localized deformations and can, therefore, better accommodate anatomical, physiological and pathological variability between patients.
- **Deformable image registration (DIR):** Deformable image registration (DIR) involves estimating the geometric transformation between two images to map them onto a common coordinate system (CCS). The process is nonlinear because the estimated transformation does not include only rigid transformations (i.e., translation and/or rotation) but also deformations (e.g., shrinking or stretching). DIR has been extensively studied for radiation therapy (RT) applications, and its integration into clinical practice is currently the object of intensive research.
- **Pixel Based Method:** For registration, a cross-correlation statistical methodology is employed in this procedure. It is frequently used for template matching or pattern recognition, which involves finding the location and orientation of a template or pattern in an image. Cross-correlation is a measure of similarity or a match metric.
- **Point Mapping Method:** Control points for point matching are crucial. Corners, intersection lines, points of locally greatest curvature on

contour lines, centres of windows with locally maximum curvature, and centres of gravity of closed-boundary areas are all control points.

- **Contour-Based Image Registration:** To produce the contour of an image, the mean for a given collection of colours is computed. Then each RGB pixel in an image is categorized as having a colour in a specific range or not during the segmentation process. In addition, the Euclidean distance is necessary to determine similarity. The locus of points is a sphere with a radius equal to the threshold value, and every point situated inside or on the sphere's surface meets the stated colour requirement. A binary, segmented image is produced by coding these two sets of points in the image with black and white. A Gaussian filter is used to eliminate noise following the segmentation process. Thresholds blurred the image, and then the contour of the image was obtained.
- **Image Registration Using Mutual Information:** Registration of multispectral/multisensory pictures is a difficult task. In general, such images have varying grey-level properties, and basic approaches based on area correlation cannot be easily used. Fortunately, it was discovered that the entropy remains constant even though the histogram changes. Even after randomly shuffling the image pixels, the entropy of the image stays constant.

In the current clinical setting, medical imaging is a vital component of many applications. Since information gained from two images in the clinical track of events is usually complementary, proper integration of valuable data obtained from the separate images is often desired. The first step in this integration process is to bring the modalities involved into spatial alignment. After registration, a fusion step is required for the integrated display of the

data involved. Anatomy correspondence with and without BT applicator for locally advanced cervical cancer needs to be quantified to merge the delivered doses of external beam radiation therapy (EBRT) and brachytherapy (BT).

### 3.3 Registration of Brachytherapy MRIs

Our dataset has pre-BT (before inserting the applicator) and at-BT images (after inserting the applicator). All the other organs and the whole anatomy are different between these two images. In this first step, we align the lumbar and pelvic spine bones, a rigid/fixed structure. To do so, we use an affine transformation, a geometric transformation that preserves collinearity and distance between points.

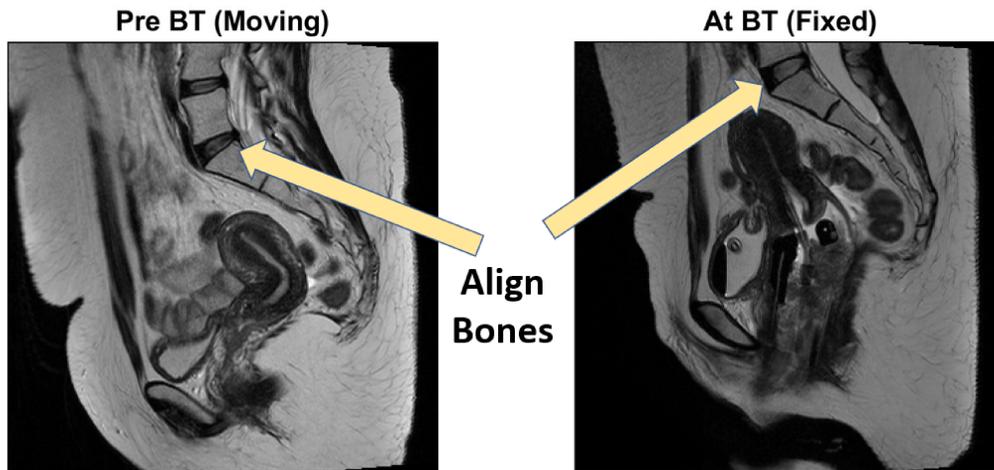


Figure 3.1: Our goal is to align the bone structure properly between pre-BT (moving) and at-BT (fixed) MRI images

## 3.4 Our Methodology

### 3.4.1 Intensity-based Image Registration

Using an intensity-based semi-automatic image registration method the pre-BT and at-BT images were aligned according to the bony anatomy (lumbar and pelvic spine). This method automatically aligns two MR images to a common coordinate system. It does not find features or uses control points. Intensity-based registration is often well-suited for medical and remotely sensed imagery. The distortion between the two images includes translation and rotation. We used an affine transformation to register the images. Optimization-based registration works best when a good initial condition can be given for the registration. In the registration process, the fixed image was the at-BT sagittal MR image, while the moving image was the pre-BT sagittal MR image. Intensity-based automatic image registration is an iterative process. It requires a pair of images, a metric, an optimizer, and a transformation type.

### 3.4.2 Affine Transformation

In the registration process, the fixed image was the at-BT sagittal MR image, while the moving image was the pre-BT sagittal MR image. The process begins with the specified transform type (affine transformation in this case) and an internally determined transformation matrix. An affine transformation is any transformation that preserves collinearity (i.e., all points lying on a line initially still lie on a line after transformation) and ratios of distances (e.g., the midpoint of a line segment remains the midpoint after transformation). Affine indicates a particular class of projective transformations that do not move objects from the affine space to the plane at infinity or conversely. In-

stead, it helps to modify the geometric structure of the image. For example, satellite imagery uses affine transformations to correct wide-angle lens distortion, panorama stitching, and image registration. Transforming and fusing the images to a large, flat coordinate system is desirable to eliminate distortion. This enables more accessible interactions and calculations that don't require accounting for image distortion. Geometric contraction, expansion, dilation, reflection, rotation, shear, similarity transformations, spiral similarities, and translation are all affine transformations, as are their combinations. In general, an affine transformation is a composition of rotations, translations, dilations, and shears.

We have chosen to perform affine transformation as we are trying to align the bone structure to preserve collinearity and distance between points. The transformation and the transformation matrix determined the specific image transformation applied to the moving image with bilinear interpolation to transform the moving image.

While an affine transformation preserves proportions on lines, it does not necessarily preserve angles or lengths. Any triangle can be transformed into any other by an affine transformation, so all triangles are affine and, in this sense, affine is a generalization of congruent and similar. A particular example combining rotation and expansion is the rotation-enlargement transformation,

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = s \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix} \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix} \quad (3.2)$$

Separating the equation,

$$x' = s(\cos\alpha)x + s(\sin\alpha)y - s(x_0\cos\alpha + y_0\sin\alpha) \quad (3.3)$$

$$y' = -s(\sin\alpha)x + s(\cos\alpha)y + s(x_0\sin\alpha - y_0\cos\alpha) \quad (3.4)$$

This can be also written as,

$$x' = ax - by + c \quad (3.5)$$

$$y' = bx + ay + d \quad (3.6)$$

where

$$a = s(\cos\alpha) \quad (3.7)$$

$$b = -s(\sin\alpha) \quad (3.8)$$

$$s = \sqrt{a^2 + b^2} \quad (3.9)$$

$$\alpha = \tan^{-1}\left(-\frac{b}{a}\right) \quad (3.10)$$

An affine transformation of  $R^n$  is a map  $F : R^n \rightarrow R^n$  of the form

$$F(p) = Ap + q \quad (3.11)$$

for all  $p$  in  $R^n$ , where  $A$  is a linear transformation of  $R^n$ . If  $\det(A) > 0$ , the transformation is orientation-preserving; if  $\det(A) < 0$ , it is orientation-reversing.

### 3.4.3 Similarity Metric

It is a common task in image analysis to compare two images' similarities. This comparison may be limited to a particular region of each image. Image Similarity Metrics are methods that produce a quantitative evaluation of the

similarity between two images or two image regions. These techniques are used as a base for registration methods because they provide the information that indicates when the registration process is going in the right direction. Many Image Similarity Metrics have been proposed in the medical image and computer vision community. There is no right image similarity metric but a set of appropriate metrics for particular applications. Metrics are probably the most critical element of a registration problem. The metric defines the goal of the process by measuring how well the reference image matches the target image after the transform has been applied. Some metrics have a rather large capture region, which means that the optimizer can find his way to a maximum even if the misalignment is high. Typically large capture regions are associated with low precision for the maximum. Other metrics can provide high accuracy for the final registration but usually require to be initialized close to the optimal value. Unfortunately, there are no clear rules about how to select a metric other than trying some of them in different conditions. In some cases, one can use a particular metric to get an initial approximation of the transformation and switch to another more sensitive metric to achieve better precision. Some well-known similarity metrics are Mutual Information (MI), Normalized Correlation (NC), Structural Similarity Index (SSIM), and Sum of Squared Differences (SSD) between intensity values. The metric used in our approach is MI. More specifically, we use Mattes Mutual Information [85]. Mutual information metrics are techniques for measuring how related two variables are. These algorithms use the joint probability distribution of a sampling of pixels from two images to measure the certainty that the values of one set of pixels map to similar values in the other image. This information is a quantitative measure of how similar the images are. High mutual information implies a large reduction in the uncertainty (entropy) between the two distributions,

signalling that the images are better aligned.

Entropy can be written as,

$$H = \sum_{i,j} -p_{ij} \log_2 p_{ij} \quad (3.12)$$

$p_{ij}$  is the probability of event  $i$  and  $j$ .

Mutual Information can be explained as,

$$I(A, B) = H(B) - H(B|A) \quad (3.13)$$

$$I(A, B) = H(A) + H(B) - H(A, B) \quad (3.14)$$

The Mattes mutual information algorithm uses a single set of pixel locations for the duration of the optimization instead of drawing a new set at each iteration. The number of samples used to compute the probability density estimates and the number of bins used to calculate the entropy are both user selectable. The marginal and joint probability density function is evaluated at the uniformly spaced bins using the samples. Entropy values are computed by summing over the bins. Zero-order and third-order B-spline kernels are used to calculate the probability density functions of the fixed and moving images, respectively.

### 3.4.4 Optimizer

The goal of image registration algorithms is to optimize the similarity measurements. The optimizer is the component that drives the execution of the registration. It is usually desirable to fine-tune the parameters of the optimizer. Each optimizer has particular parameters that must be interpreted in the context of its implementation optimization strategy. The optimizer used

in this example is a variant of gradient descent that attempts to prevent it from taking too large steps. At each iteration, this optimizer will take a step along the direction of the derivative.

In this case, the optimizer (gradient descent) defines the methodology for optimizing the similarity metric. The registration is improved by adjusting the optimizer and metric configuration properties. The *Initial-Radius* property of the optimizer controls the initial step size used in parameter space to refine the geometric transformation. The parameter is fine-tuned to obtain a noticeable improvement in the alignment of the rigid/bone structure in the images. Along with the bone structure, the position of the uterus was also aligned, which helps to measure the uterus deformation more accurately.

Finally, the optimizer checks for a stop condition. A stop condition is anything that warrants the termination of the process. In most cases, the process stops when it reaches a point of diminishing returns or when it reaches the specified maximum number of iterations. If there is no stop condition, the optimizer adjusts the transformation matrix to begin the next iteration.

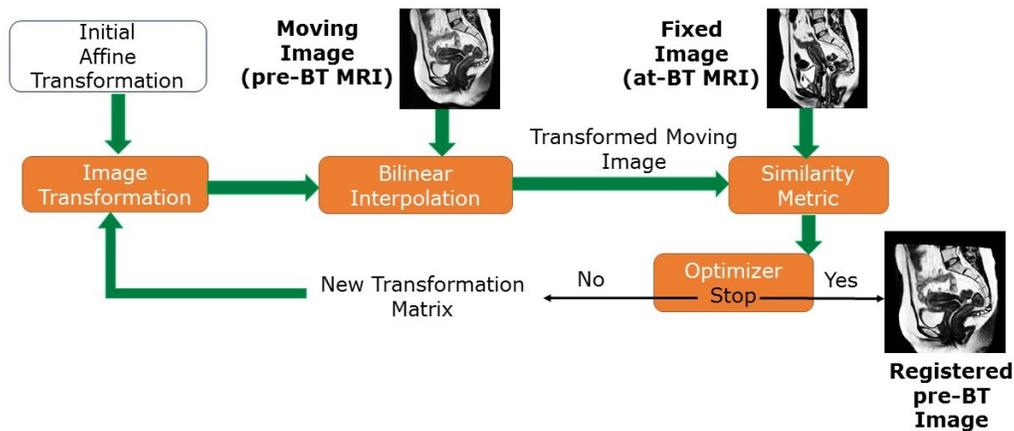


Figure 3.2: Automated image registration algorithm to align spine between pre-BT and at-BT MRI

## 3.5 Comparison with Other Methods

### 3.5.1 Optical Flow

Optical flow is a method for motion analysis and image registration that aims to compute displacement of intensity patterns. Optical flow is used in many different settings in the computer vision problems, such as video recognition and video compression. Optical flow methods are based on computing estimates of the motion of the image intensities over time. This method can be used to detect and delineate independently moving objects, even in the presence of camera motion. Of course, optical-flow-based techniques are computationally complex, and hence require fast hardware and software solutions to implement. Since optical flow is fundamentally a differential quantity, estimation of it is highly susceptible to noise. The motion of human bodies can also be analyzed with optical flow. There are many methods of estimating the optical flow between two frames, including differential-based, region-based, energy-based, and phase-based methods [86]. Large displacement is a fundamental area of concern, responsible for the failure of many optical flow algorithms. It occurs as a result of motion of an object moving at a high speed or due to a low frame-rate. The majority of the algorithms fail to tackle with large displacement because the energy function may be trapped into an incorrect local minimum. In our dataset, we also have very large organ displacements and variations in patient anatomy. For those cases optical flow based algorithms were not able to perform very good compared to our proposed image registration method.

### 3.5.2 Elastix

Elastix is an image registration toolbox built upon the Insight Segmentation and Registration Toolkit (ITK) [87]. It is entirely open-source and provides a wide range of algorithms employed in image registration problems. Its components are designed to be modular to ease a fast and reliable creation of various registration pipelines tailored for case-specific applications. The software consists of a collection of algorithms that are commonly used to solve medical image registration problems. The modular design of elastix allows the user to quickly configure, test, and compare different registration methods for a specific application. The command-line interface enables automated processing of large numbers of data sets, by means of scripting. The usage of elastix for comparing different registration methods is illustrated with three example experiments, in which individual components of the registration method are varied.

## 3.6 Results

Table 3.1: Evaluation of Intensity-based image registration method. **Structural Similarity Index Measure (SSIM)** between two images are reported. SSIM between fixed image and registered image, fixed image and moving image are reported.

A value closer to 1 indicates better image similarity.

	<b>Fixed &amp; Registered</b>	<b>Fixed &amp; Moving</b>
<b>Intensity-based</b>	<b>0.45</b>	<b>0.25</b>
Optical Flow	0.25	0.25
Elastix	0.16	0.25

Table 3.1 displays Structural Similarity Index Measure (SSIM) between two images. SSIM between fixed image and registered image, fixed image and moving image are reported here. From the results, we can conclude that our

Table 3.2: Evaluation of Intensity-based image registration method. **Mutual Information (MI)** between two images are reported. MI between fixed image and registered image, fixed image and moving image are reported. The larger the value, the greater the relationship between the two images.

	<b>Fixed &amp; Registered</b>	<b>Fixed &amp; Moving</b>
<b>Intensity-based</b>	<b>0.75</b>	<b>0.35</b>
Optical Flow	0.45	0.35
Elastix	0.36	0.35

method, i.e. Intensity-based Image Registration out-performed all the other methods.

In Table 3.2, Mutual Information (MI) between two images are displayed. For example, MI between fixed and registered images, and fixed and moving images are reported. The larger the value, the greater the relationship between the two images.

In this section, the visual results of the intensity-based image registration method are shown in Fig. 3.3, 3.4, 3.5, 3.6. For each of these examples, one can observe that the alignment of the bone structure (lumbar and spline) is different between the moving image (pre-BT) and fixed image (at-BT). As the only rigid/fixed part (point of reference) between these two image pairs is the bone structure, we had to align the bone so that the other organs, especially the uterus, are correctly aligned for further experiments. To align the bone structure, the moving image (pre-BT) is transformed (translation and rotation) vertically and horizontally to match the fixed image (at-BT). As there is large deformation in the anatomy for every patient between pre-BT and at-BT images, the problem of registering these images is challenging. Nevertheless, this automated registration method worked well in aligning the fixed anatomy.

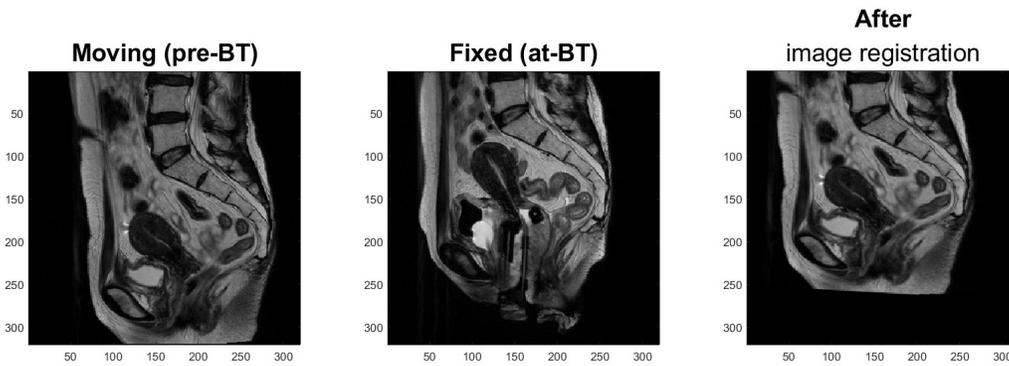


Figure 3.3: To align the bone structure, the moving image is transformed vertically and horizontally to align with the fixed image

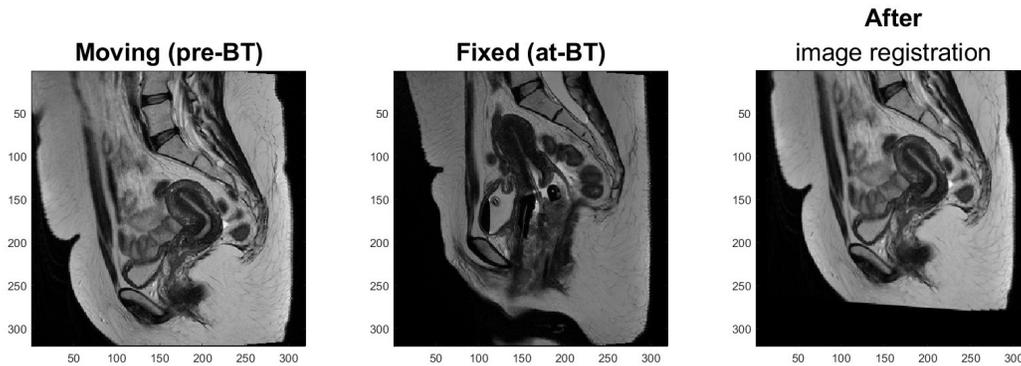


Figure 3.4: Another example of the registration method where the organs change shape and position drastically

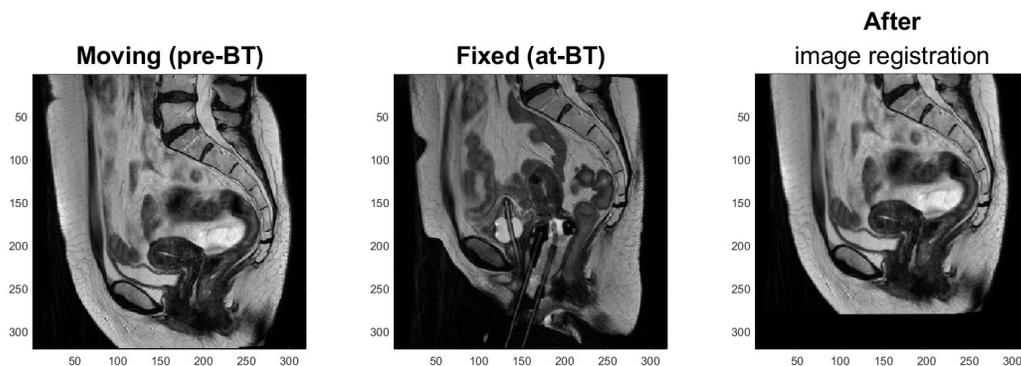


Figure 3.5: There is a large uterine deformation in this example which makes the registration problem harder to solve as the only point of reference is the bony anatomy

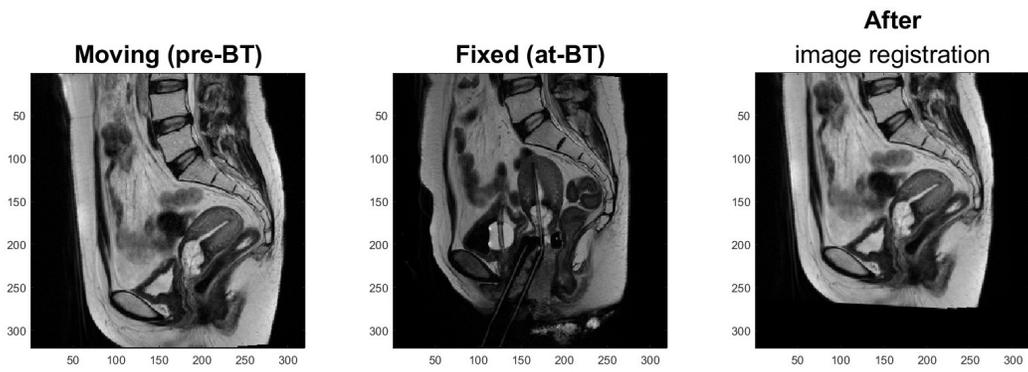


Figure 3.6: Another example of the registration method where the uterus is retroverted

## 3.7 Discussion

From the results shown in Table 3.1 and Table 3.2, we can conclude that in both measures, SSIM and MI, our method performed better than the other reported methods. Although registering only the bone structures between pre-BT and at-BT images is a very challenging task. A value closer to 1 indicates better image similarity. The larger the value, the greater the relationship between the two images. The SSIM between the fixed image and the moving image is 0.25. After registering the moving image to the fixed image, the SSIM between the fixed and registered image is 0.45. As the other organs in these images are very different and there is a drastic change in the other anatomy apart from the bone structure, the SSIM cannot be higher than the values we obtained. This registration method helped in aligning the position of the bone, as well as the uterus. This step is very useful for further computations as we needed to align the uterus properly to predict the uterus deformation from pre-BT to at-BT position.

# Chapter 4

## Automated Uterus

## Segmentation Using CNN

### 4.1 Overview

Medical image segmentation is essential in computer-aided diagnosis systems in different applications. The vast investment and development of medical imaging modalities such as microscopy, X-ray, ultrasound, computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography attract researchers to implement new medical image-processing algorithms. Medical image segmentation involves the extraction of regions of interest (ROIs) from image data. The main goal of segmenting this data is to identify areas of the anatomy required for a particular study. Medical image segmentation can be time-consuming, and recent advances in ML software techniques are making it easier for routine tasks to be completed. One of the key benefits of medical image segmentation is that it allows for a more precise analysis of anatomical data by isolating only necessary areas. For specific procedures, such as implant design, it is essential to segment specific struc-

tures, such as the hip or knee. In addition, segmentation offers the benefit of removing any unwanted details from a scan, such as air and allowing different tissues, such as bone and soft tissues, to be isolated. When combined with different software processing options, researchers and clinicians can generate a series of segmented masks ready for further analysis. Medical image segmentation has automatic or semiautomatic detection of two-dimensional (2D) or three-dimensional (3D) images.

## 4.2 Various Medical Image Segmentation

### Methods

Deep learning has become the mainstream of medical image segmentation methods. In particular, for automatic medical image segmentation with different modalities of images (MRI, CT), great advances have been made using various DL models. Image segmentation aims to simplify or change the representation of images, making them easier to understand and analyze. The emergence of DL has made it possible to segment medical images efficiently and effectively, even for small datasets. Segmentation algorithms based on convolutional neural networks (CNNs) have already become the standard approach in image segmentation tasks. Their excellent segmentation ability has been demonstrated experimentally and theoretically and can be further applied to medical images. The majority of segmentation methods follow the structure of the FCN [88] networks and U-Net [89]. The Fully Convolutional Network (FCN) has been increasingly used in different medical image segmentation problems. The FCN was introduced in the image segmentation domain as an alternative to image patches. Using FCN, the image could be analyzed globally instead of using localized patches. Thus, there is no need to select rep-

representative patches, eliminate redundant calculations where patches overlap, and scale up more efficiently with image resolution. Moreover, there can be a fusion of different scales by adding links that combine the final prediction layer with lower layers with finer strides. Since convolution layers replace all the fully connected layers, the FCN can take inputs of arbitrary sizes and produce correspondingly-sized outputs with efficient inference and learning. Unlike patch-based methods, this architecture's loss function is computed over the entire image segmentation result. The most common FCN used in medical imaging applications is the U-Net [89]. The U-Net architecture consists of a contracting path to capture context and a symmetric expanding path that enables precise localization. In general, the FCN architectures can be used for tasks involving pixel-wise loss functions such as reconstruction, synthesis, and other image-to-image tasks.

In our problem, we utilized the CNNs to automatically segment the region of interest, i.e. the uterus from the MRI images. Manual segmentation by an expert radiologist is the widely accepted ground truth/gold standard. But this manual segmentation is a very time-consuming, tedious task, usually involving lengthy procedures, and the results depend on human expertise. Moreover, these results vary from expert to expert and generally are not reproducible by the same expert. Thus automatic segmentation and reproducible segmentation methods are very much in demand nowadays.

We proposed an automated segmentation approach to delineate uterus structure on MRI using a pre-trained deep convolutional neural network (CNN) based Inception V4 augmented with a multi-layer autoencoder. A detailed step-by-step explanation is given below.

## 4.3 Uterus Segmentation using Deep CNN

### Inception-V4

#### 4.3.1 Proposed Deep ConvNet Architecture

Here we utilized the Inception-V4 [90] pre-trained on ImageNet Large Scale Visual Recognition Competition (ILSVRC) datasets and updated the network parameters for our MRI dataset. The original Inception V4 network architecture is shown in Fig. 4.1, which we fine-tuned according to our problem formulation. The first convolutional layer and the last fully connected layers are modified by changing the initial weights according to our problem. The Inception-V4 was trained on RGB colour images with three channels different from the MRI dataset used in this study consisting of grey-scale values. Therefore, the three channels of CNN were reduced to 1 channel. The ImageNet version of the algorithms comes with a softmax layer as the last layer with 1000 categories. However, it does not include an approach to predict the object boundaries; therefore, the task of predicting segmentation contour (i.e., 400x1 vector) should be developed for the proposed problem. As such, the pre-trained network's last layer (softmax layer) is truncated and replaced with a new softmax layer with a dimension of 400. In the proposed neural net approach, the mean squared error (MSE) is used as the loss function, and the Adam version of the stochastic gradient descent is used as the optimizer [91].

#### 4.3.2 Utilizing Transfer Learning

Transfer learning is a highly popular technique nowadays in which a DL network trained on a large dataset from one domain is used to retrain or fine-tune the DL network with a smaller dataset associated with another domain. How-

ever, the limited size of the annotated medical image datasets and the current trend of using deeper and larger structures increase the risk of overtraining and make transfer learning more appealing in medical imaging. In transfer learning, the neural network is trained in two stages: 1) Pre-training, where the network is generally trained on a large-scale benchmark dataset representing a wide diversity of labels/categories (e.g., ImageNet); and 2) Fine-tuning, where the pre-trained network is further trained on the specific target task of interest, which may have fewer labelled examples than the pre-training dataset. The pre-training step helps the network learn general features that can be reused on the target task.

This two-stage paradigm has become extremely popular in many settings, particularly in medical imaging. In the context of transfer learning, standard architectures designed for ImageNet with corresponding pre-trained weights are fine-tuned on medical tasks.

### **4.3.3 Dimensionality Reduction Using Autoencoders**

Autoencoders and their deep version are traditionally dimensionality reduction methods that have achieved great success via the powerful representability of neural networks. Autoencoders belong to a particular family of dimensionality reduction methods implemented using ANN. It aims to learn a compressed representation for input by minimizing its reconstruction error [92]. Furthermore, it demonstrates a good ability to learn meaningful features from data [92]. An autoencoder generally consists of two parts: an encoder that transforms the input to a hidden code and a decoder that reconstructs the input from hidden code.

As soon as the image is reconstructed, you compare the reconstructed image with the original image, compute the difference, and calculate the loss

which can then be minimized.

The Loss is calculated by:

$$L(\theta, \phi) = \frac{1}{n} \sum_i (x^i - f_{\theta}(g_{\phi}(x^i)))^2 \quad (4.1)$$

The hidden layer size is smaller than the input layer in autoencoders and is 50 in our case. By reducing the hidden layer size, we force the network to learn the important features of the dataset. The main reason for using autoencoders in our problem is to reduce the dimensionality and reconstruct the original contour from the hidden representation. The non-linearity property of the autoencoders captures the spatial information regarding the shape and sharp edges of the segmentation contour. Before computing the hidden representation, we impose the non-linear activation function, rectified linear unit (ReLU), to introduce non-linearities in the linear output.

In our implementation, at first, we used a multi-layer autoencoder. We trained it with the segmentation contour, i.e. the 400x1 vector (200 x-coordinates and 200 y-coordinates), and the bottleneck dimension is 50x1. Then the CNN is trained with the target as the bottleneck vector of the autoencoder (i.e. 50x1 latent representation). After CNN's prediction, we attach the decoder part of the autoencoder. CNN's prediction (i.e. the 50x1 vector) is passed through the decoder to reconstruct the original 400x1 vector, which produces the segmentation contour. We train both the models end-to-end as shown in Figure 4.2. In the autoencoders, we impose the non-linear activation function ReLU to capture complicated properties such as shape and size related to the underlying segmented structure.

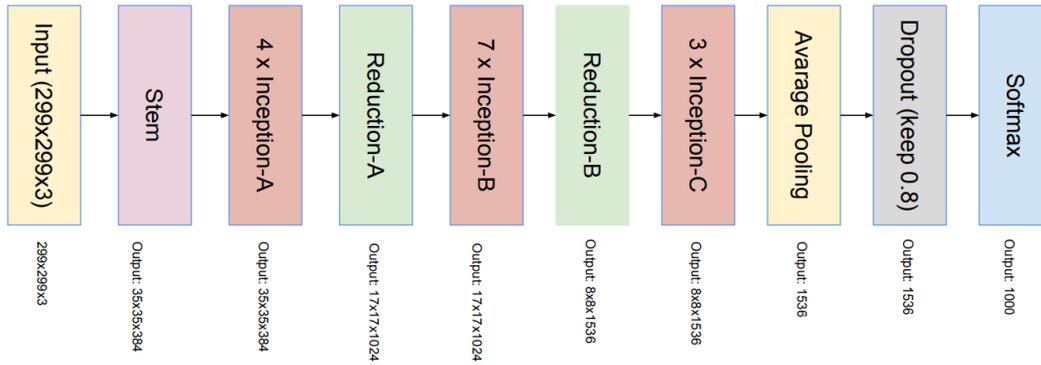


Figure 4.1: The schema for the pure Inception-V4 network

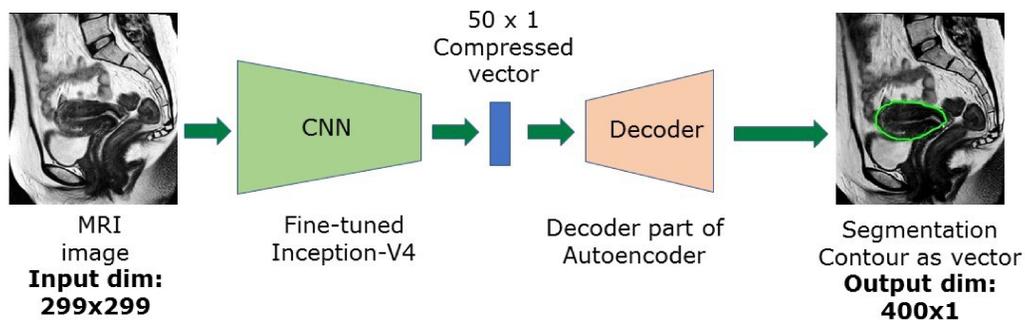


Figure 4.2: Components of proposed neural net approach which consists of fine-tuned CNN architecture combining decoder part of auto-encoder

## 4.4 Uterus Segmentation using Deep U-Net

### 4.4.1 U-net Architecture

The application of deep learning technology in medical imaging has attracted extensive attention. How to automatically recognize and segment the lesions in medical images has become an issue that concerns many researchers. Ronneberger et al. [89] proposed U-Net at the MICCAI conference in 2015 to tackle this problem, a breakthrough of deep learning in medical imaging segmentation. U-Net is a Fully Convolutional Network (FCN) applied to biomedical image segmentation, composed of the encoder, the bottleneck module, and the decoder. The widely used U-Net meets the requirements of medical image segmentation for its U-shaped structure combined with context information, fast training speed, and a small amount of data used. The original U-Net architecture is illustrated in Fig. 4.3. We compared our proposed method (Inception-V4 with autoencoders) with U-net for the segmentation task. The proposed network uses four depth levels with two convolutional layers in each depth level and the bridge of the network. In our implementation, we used 32, 64, 128, 256, and 512 channels in the feature maps at levels 1, 2, 3, 4, and the bridge, respectively. This architecture consists of three sections: contraction, bottleneck, and expansion. Each block in contraction takes an input and applies two  $3 \times 3$  convolution layers, followed by a  $2 \times 2$  max pooling. The number of kernels or feature maps after each block doubles so that the architecture can learn the complex structures effectively. The bottom-most layer mediates between the contraction layer and the expansion layer. It uses two  $3 \times 3$  CNN layers followed by a  $2 \times 2$  up-convolution layer.

We compared the proposed method with the current state-of-the-art method:

U-net. The network architecture without input/output MRI images is illustrated in Fig. 4.4. The U-Net combines the location information from the down-sampling path with the contextual information in the up-sampling path to obtain general information combining localisation and context, which is necessary to predict a good segmentation map.

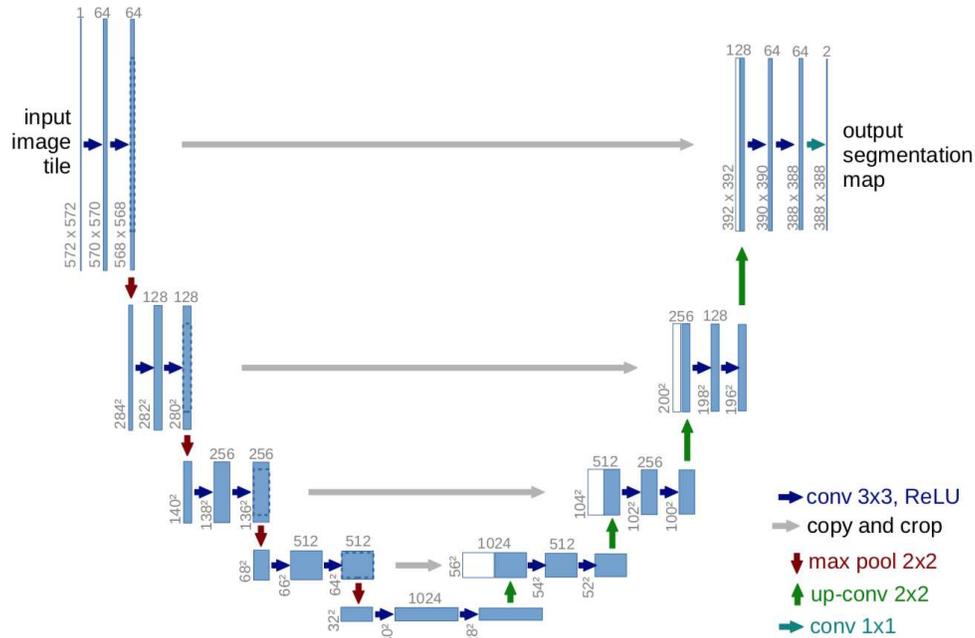


Figure 4.3: Illustration of U-Net convolution network structure. The left side of the U-shape is the encoding stage, also called the contraction path and the right side of the U-shape, also called the expansion part, consists of the decoding stage and the up-sampling process

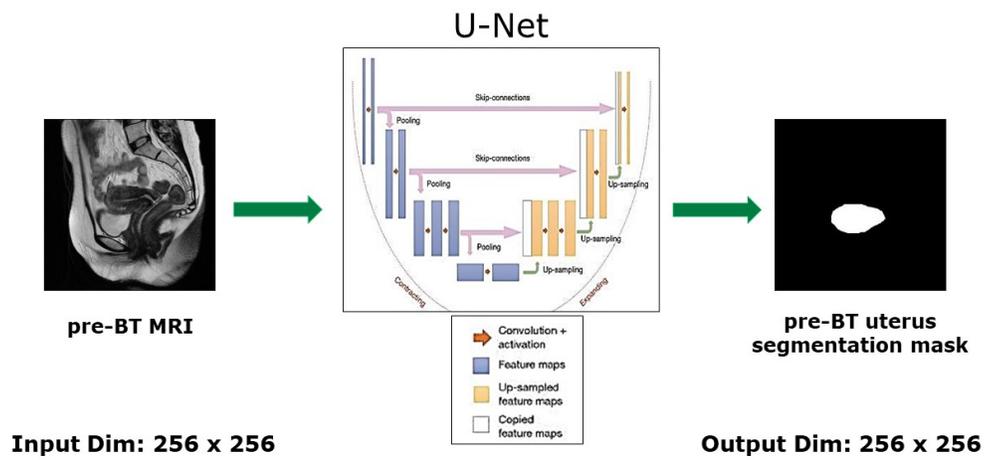


Figure 4.4: Illustration of U-Net convolution network structure for our problem formulation

## 4.5 Results

### 4.5.1 Data Preprocessing

Step 1: Originally, all the MR images and the corresponding contours were in a DICOM format in the standard DICOM patient-based coordinate system. We had to convert the DICOM voxel to the patient coordinate system. In conjunction with the Pixel Spacing Attribute, the Image Plane Attributes describe the position and orientation of the image slices relative to the patient-based coordinate system. The Image Position (Patient) specifies the image's origin concerning the patient-based coordinate system in each image frame. Image Orientation (Patient) attribute values specify the orientation of the image frame rows and columns. The mapping of pixel location (i, j) is calculated as shown in Figure 4.5 [93]:

$$\begin{bmatrix} P_x \\ P_y \\ P_z \\ 1 \end{bmatrix} = \begin{bmatrix} X_x \Delta i & Y_x \Delta j & 0 & S_x \\ X_y \Delta i & Y_y \Delta j & 0 & S_y \\ X_z \Delta i & Y_z \Delta j & 0 & S_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} i \\ j \\ 0 \\ 1 \end{bmatrix} = \mathbf{M} \begin{bmatrix} i \\ j \\ 0 \\ 1 \end{bmatrix}$$

Where:

- $P_{xyz}$  The coordinates of the voxel (i,j) in the frame's image plane in units of mm.
- $S_{xyz}$  The three values of the Image Position (Patient) (0020,0032) attributes. It is the location in mm from the origin of the RCS.
- $X_{xyz}$  The values from the row (X) direction cosine of the Image Orientation (Patient) (0020,0037) attribute.
- $Y_{xyz}$  The values from the column (Y) direction cosine of the Image Orientation (Patient) (0020,0037) attribute.
- $i$  Column index to the image plane. The first column is index zero.
- $\Delta i$  Column pixel resolution of the Pixel Spacing (0028,0030) attribute in units of mm.
- $j$  Row index to the image plane. The first row index is zero.
- $\Delta j$  Row pixel resolution of the Pixel Spacing (0028,0030) attribute in units of mm.

Figure 4.5: Transformation from DICOM to patient-coordinate system

Step 2: The number of corresponding contour points in the pre-BT image and

at-BT image were different. Therefore, we performed a uniform sampling of contour points to generate the same number for each dataset. Each contour is sampled to 200 points using the spline interpolation, and the sampling is performed by transforming the contours to the polar coordinate system [94].

Step 3: Originally each 2D MR image was of size 320x320 pixels. Therefore, the images must be resized to fit the input shape of the network. The input to the proposed network is a 2D MR image of size 299x299, and the output is the segmentation contour, i.e., the contour's stacked (x; y) coordinate points. We used 200 pairs of (x; y) points and stacked them to form a vector of size 400x1. Each input image and the segmentation mask for U-net is resized to 256x256.

## 4.5.2 Data Augmentation

Data augmentation creates new samples based on existing samples in a dataset or according to a generative model. These new samples can then be combined with the original samples to increase the variability in data points in a dataset. This class of techniques has become a common practice in DL-based applications since it is highly effective for increasing the size of training sets, reducing the chance of overfitting and eliminating the unbalance issue in datasets is critical for achieving generalized models and testing results.

Common data augmentation techniques adopted in medical image analysis applications [89] include cropping, translation, rotation, flipping, and scaling of images. Here we rotated the MR images with different degrees (5, 10, 15, 20, 25, 30) clockwise and anti-clockwise to generate more training examples to train the DL models. At the same time, the same operation applied to an MR image is used for corresponding mask images. In this way, each aug-

mented image pair’s relationship between the MR image and the output mask is preserved.

### 4.5.3 Implementation Details

We input a 299x299x1 grayscale MRI image into the InceptionV4 to predict a 50x1 response vector using an autoencoder. The training was performed using 250 epochs and a batch size of 6. The last layer is a dense layer that uses the MSE between the actual and the predicted value as the loss function and is illustrated in Eq. 4.2. Adam’s version of the stochastic gradient descent [91] is used as the optimizer. All the experiments with the neural network were implemented using Keras API [95] with TensorFlow and on two NVIDIA GeForce GTX 1080 GPU processors. It took around 5 hours to train the network. The prediction took 4 ms per image.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4.2)$$

### 4.5.4 Quantitative Evaluation Metrics

For evaluation, we used Dice Coefficient (DC), and the Hausdorff Distance (HD) between the ground truth segmentation and the segmentation obtained using the automated methods. Also, RMSE is used to measure the difference between the source and segmented images. The smaller the value of RMSE, the better the segmentation performance. A higher DC value is associated with a better segmentation result. The lower value of Hausdorff distance corresponds to better segmentation.

### 4.5.5 Uterus Segmentation Results -

#### Our Proposed Method vs U-Net

Table 4.1 reports the DC and the HD between manual segmentation and the segmentation by automated methods. The proposed CNN with autoencoder performed significantly better than the current state-of-the-art method U-net in terms of Hausdorff distance and dice score.

Table 4.1: Evaluation of automated segmentation results by the proposed method and U-net in terms of Dice Coefficient (DC) and Hausdorff Distance (HD) in comparison to expert manual segmentation. The Mean  $\pm$  Standard Deviation values are reported for each metric. The higher the Dice coefficient or the lower the Hausdorff distance the better the results.

	Dice Coefficient (%)	Hausdorff Distance (mm)
<b>Our Method</b>	<b>94.8 <math>\pm</math> 2.3</b>	<b>3.06 <math>\pm</math> 2.5</b>
<b>U-Net</b>	92.4 $\pm$ 4.0	6.7 $\pm$ 3.6

The MSE loss curves (training and validation curves) with different epochs for CNN with autoencoders are shown in Figure 4.6.

In Figure 4.7, the highest accuracy (97%), 50th percentile (93%), 25th percentile (90%) and the lowest accuracy (86%) of Dice scores achieved by the proposed method are shown visually, where the green, yellow and red contours depict the segmentation results by the proposed method (CNN with autoencoder), U-net approach and manual ground truth, respectively.

**Here are some unusual results that are produced by our DL method and U-net.**

## 4.6 Discussion

Deep Learning has pushed the limits of what was possible in Image processing. DL methods such as CNN mainly improve prediction performance using

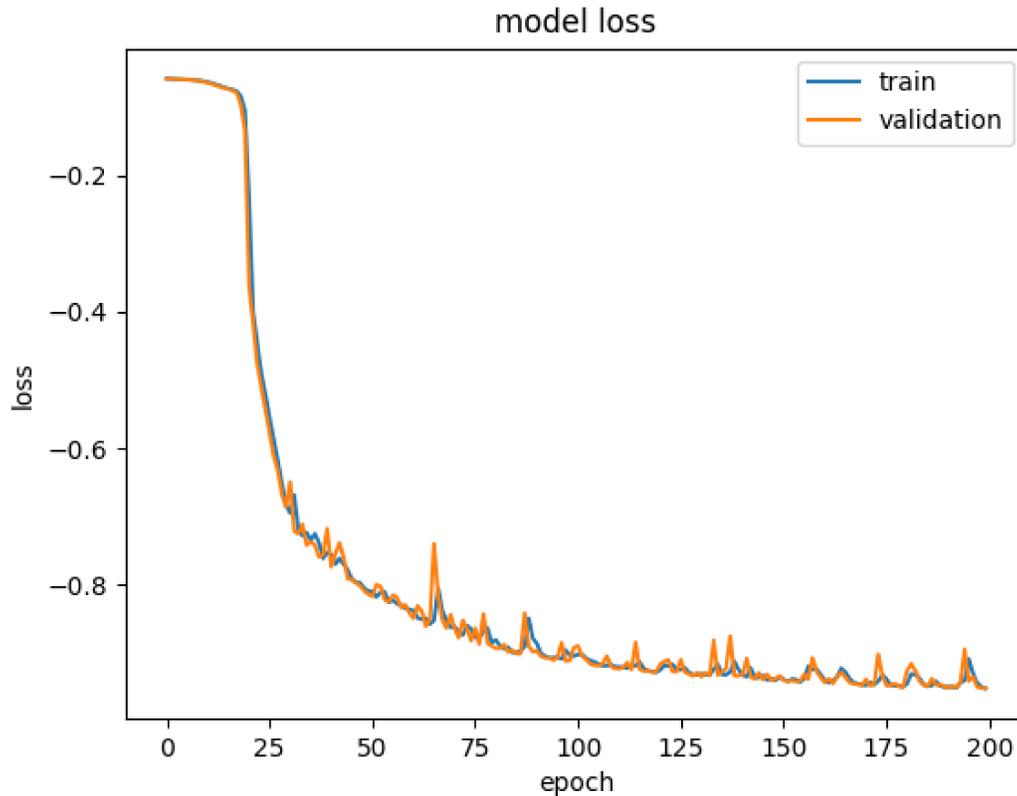


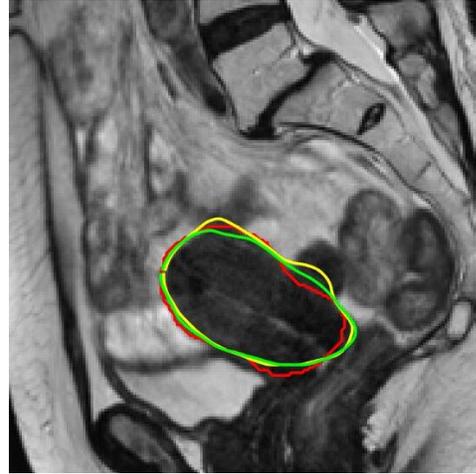
Figure 4.6: Loss - (MSE) vs. Epoch curve for training and validation for CNN with autoencoder

big data and plentiful computing resources and have pushed the boundaries of what was possible. Problems assumed to be unsolvable are now being solved with super-human accuracy. DL has dominated the domain due to a substantially better performance than traditional methods. Also, DL is not going to solve all CV problems. There are problems where traditional techniques with global features are a better solution. Nevertheless, the advent of DL may open many doors to do something with traditional techniques to overcome the challenges DL brings (e.g. computing power, time, accuracy, characteristics and quantity of inputs, among others).

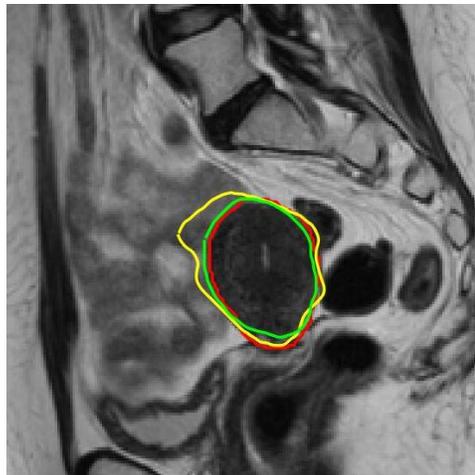
The difficulty with the traditional approach is that it is necessary to choose which features are essential in each given image. As the number of classes to



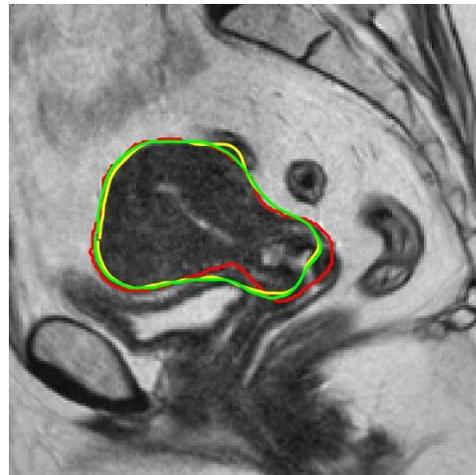
(a) Highest DC: CNN - 0.97 [U-Net - 0.94]



(b) CNN - 0.95 [U-Net - 0.92]



(c) CNN - 0.93 [U-Net - 0.88]



(d) CNN - 0.91 [U-Net - 0.92]



(e) CNN - 0.90 [U-Net - 0.88]



(f) Lowest DC: CNN - 0.85 [U-Net - 0.84]

Figure 4.7: The automated segmentation results of CNN with autoencoders and U-net (in square brackets) with different Dice score (DC) values. The red, green, and yellow contours represent the ground truth, segmentation by CNN with autoencoders and U-net, respectively.

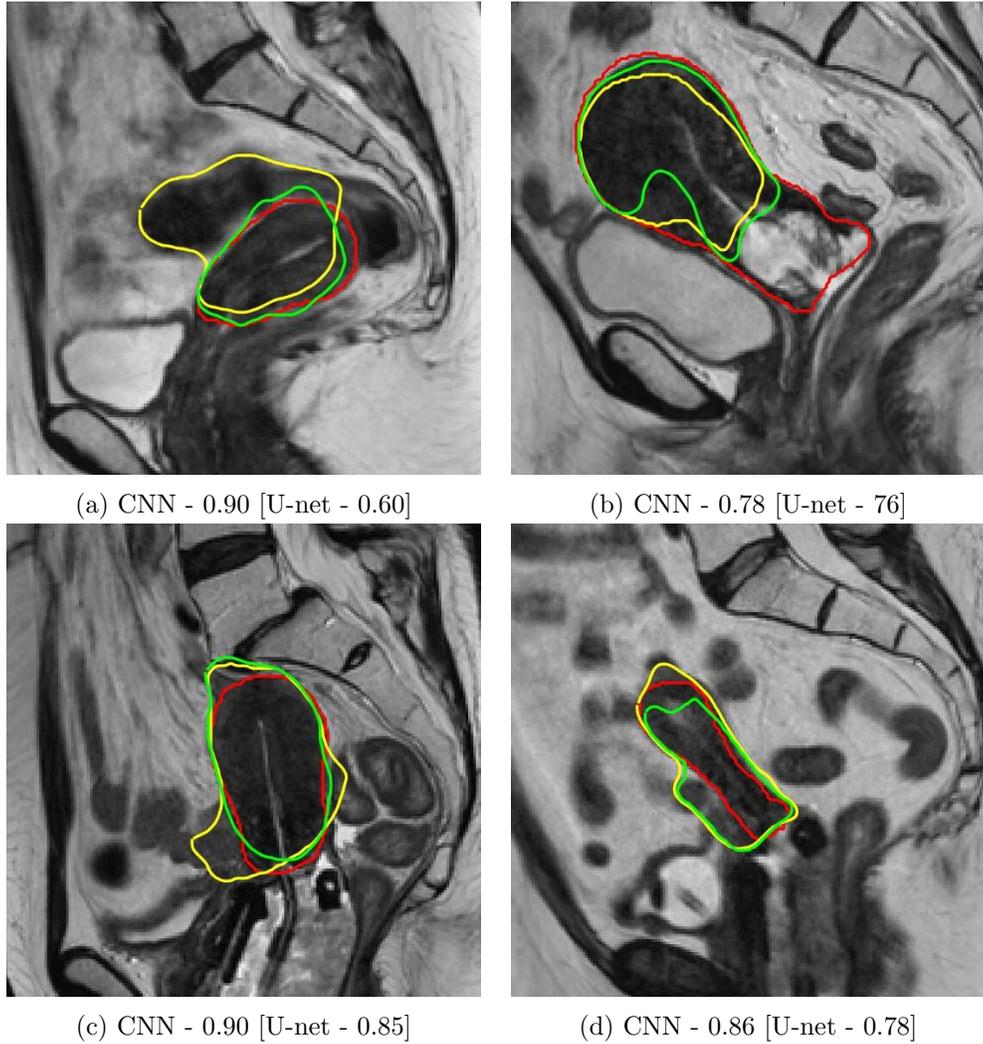


Figure 4.8: Unusual results predicted by DL methods. Segmentation results of CNN with autoencoders and U-net (in square brackets) with different Dice score (DC) values. The red, green, and yellow contours represent the ground truth, segmentation by CNN with autoencoders and U-net, respectively.

classify increases, feature extraction becomes more and more cumbersome. It is up to the /clinician’s judgment and a lengthy trial and error process to decide which features best describe different classes of objects. Moreover, each feature definition requires dealing with many parameters, all of which must be fine-tuned manually.

DL introduced the concept of end-to-end learning, where the machine is just given a dataset of images which have been annotated with labels. A DL model is ‘trained’ on the given data, where neural networks discover the underlying patterns in the data and automatically work out the most descriptive and salient features with respect to each specific class of object. It has been well-established that deep neural networks perform far better than traditional algorithms, albeit with trade-offs concerning computing requirements and training time.

For our dataset, CNN with autoencoder performed better than U-net. The proposed approach incorporates an autoencoder to detect the segmentation contours directly from the raw MR images. Previous unsupervised approaches, such as principal component analysis-based linear reconstruction, are not suitable [26] since they cannot preserve the highly nonlinear relationships when projected to low-dimensional space. Therefore, we imposed a robust structured regression approach to the proposed model by applying an autoencoder. The proposed automated segmentation method can be utilized to segment the closed and open regions like atrial segmentation [26]. This unique segmentation approach can be easily generalized from 2D to 3D images. Also, it can be used in most segmentation tasks as this method does not depend on any specific condition.

Most of the existing machine learning-based methods, such as U-net, rely on semantic segmentation, which may not be suitable for all the image seg-

mentation tasks like predicting the open contours. After indicating the closed contour, many post-processing steps must be done to extract the available contour. But with our proposed method, we can directly predict any segmentation contour in spite of having different shapes and orientations. U-Net is one of the fully connected networks that outperformed the classical image processing approaches in many applications, such as liver segmentation, brain tumour segmentation, prostate segmentation etc. However, U-Net is limited in extracting some complex features that could help image segmentation in medical images. For example, U-Net can detect objects of known shape and location but is limited in extracting complex features when the target object has a non-standard shape and random location.

So, with our proposed uterus segmentation method, we could achieve 94.8% accuracy in terms of Dice Coefficient, and the highest DC was 97% which is shown in Figure 4.7, whereas U-Net achieved 92% accuracy in terms of DC and the highest was 94%. But an exciting finding from this step is the unusual cases shown in Figure 4.8. One can see that U-Net segmented some incorrect regions along with the correct segmentation (i.e. the uterus). So, the areas with similar intensity values as the uterus on the MRI images are included in the segmentation result, which is unacceptable in this case. That is why, with our dataset, our proposed segmentation approach CNN with autoencoder is proved to be the better choice over the U-Net to segment uterus structure automatically from MRI.

# Chapter 5

## Uterus Structure Prediction Using Deep Learning

### 5.1 Overview

Deep learning methods are highly effective when the number of available samples is large during a training stage. For example, in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), more than 1 million annotated images were provided [96]. However, as for medical applications, we usually have a very limited number of images. Therefore, one of the main challenges in applying deep learning to medical images arises from the limited number of available training samples to build deep models without suffering from overfitting. To this end, research groups have devised various strategies, such as (i) to take image patches either 2D or 3D as input [97–99], rather than the full-sized images, to reduce the input dimensionality, thus the number of model parameters; (ii) to expand their dataset by artificially generating samples via various data augmentation methods (like translation, rotation, transformation, scaling) and then train their network from scratch with the augmented

dataset [100, 101]; (iii) to use deep models trained over a huge number of natural images in computer vision as feature extractor and then train the final classifier or output layer with the target-task samples [99]; (iv) to initialize model parameters with those of pre-trained models from non-medical or natural images and then fine-tune the network parameters with the task-related samples [100]; (v) to use models trained with small-sized inputs for arbitrarily-sized inputs by transforming weights in the fully connected layers into convolutional kernels [102].

In terms of the input types, we can categorize deep models as typical multi-layer neural networks that take input values in vector form (i.e., non-structured) and convolutional networks that take 2D or 3D shaped (i.e., structured) values as input. Because of the structural characteristic of images (i.e., the structural information among neighbouring pixels or voxels is another critical source of data), convolutional neural networks have gained significant interest in medical image analysis. However, networks with vectorized inputs were also successfully applied to different medical applications [103, 104].

The numerous development in medical imaging acquisition systems and deep learning technologies have resulted in the rise of usage frequency of modalities for computer-aided diagnosis. Despite the U-Net being super-efficient in the semantic segmentation task, there is still a void to fill with improvements in certain aspects. The most apparent problem in the original (vanilla) U-Net is that the learning may slow down in deeper layers of the U-Net model, which increases the possibility of the network ignoring the layers representing abstract features of the target structure. Therefore, researchers modify the U-net according to the specifications of the problem.

The driving focus of the medical imaging community has been on the supervised learning of decision boundaries, while generative tasks were the sec-

ondary position or less important. This changed dramatically with the advent of Generative Adversarial Networks (GANs) [105], which lead to a new age of generative modeling and distribution learning. With their abilities to mimic data distributions and to synthesize images at yet unprecedented levels of realism, GANs opened new ways to bridge the gap between supervised learning and image generation. The mentioned improvements are essentially due to the following properties: 1) GANs maximize the probability density over the data-generating distribution by exploiting density ratio estimation (Isola et al., 2017) in an indirect fashion of supervision; 2) GANs can discover the high dimensional latent distribution of data, which has lead to significant performance gains in the extraction of visual features.

In this thesis, we are dealing with a unique problem in medical imaging, i.e. predicting the anatomy deformation using MRI, so the main goal is to predict the at-BT MRI (after inserting the applicator) from pre-BT MRI (before inserting the applicator). We can conclude from the images shown before that it is very difficult to address the challenging task of anatomy prediction using only the 2D MRI slices. Also, image to image prediction is a tough problem to solve. There are well-known methods for predicting an image mask (segmentation) like U-Net, V-Net, variations of U-Net architectures, fully connected networks etc. The GANs are used to generate synthetic images from real images. The architecture consists of a generator model for outputting new plausible synthetic images and a discriminator model that classifies images as real (from the dataset) or fake (generated). That means there should be some similarity between the real Image and the predicted Image. But in our dataset, there is no such similarity between pre-BT and at-BT MRI images, and the problem is to predict the anatomical deformation due to applicator insertion. For this reason, we have chosen U-Net architecture and utilized a modified

version of U-Net to predict the at-BT uterus shape and position from the pre-BT MRI and the corresponding pre-BT uterus segmentation. The fine-tuning details are described below.

## 5.2 U-Net Architecture

First introduced in 2015 in the paper [89], the U-Net possesses an approach to image segmentation that outperformed its competitors at the time, a sliding window convolutional network, all the while using few images in the training dataset and making use of image augmentation to increase the learning capability of the network. The U-Net is a CNN architecture design for image segmentation or detection applications [89]. Our framework’s prediction network is designed based on a modified version of the original U-net architecture proposed in [89]. U-shaped network consisting of a contracting path and an expansive path. The basic intuition is that while on the downslope (contracting path), the network learns to classify the object, and on the upslope (Expansive path), the networks on the localization of the object. U-Nets consist of convolution layer, batch normalization [106], rectified linear unit (ReLU), and contracting path connection with concatenation. The network uses four depth levels with two convolutional layers in each depth level and the bridge of the network. In our implementation, we used 32, 64, 128, 256, and 512 channels in the feature maps at levels 1, 2, 3, 4, and the bridge, respectively. This architecture consists of three sections: contraction, bottleneck, and expansion. Each block in contraction takes an input and applies two  $3 \times 3$  convolution layers, followed by a  $2 \times 2$  max pooling. The bottom-most layer mediates between the contraction layer and the expansion layer. It uses two  $3 \times 3$  CNN layers followed by a  $2 \times 2$  up-convolution layer. Specifically, each stage contains four

sequential layers composed of convolution with  $3 \times 3$  kernels, batch normalization, and ReLU layers. Finally, the last scene has two sequential layers, and the previous layer contains only the convolution layer with a  $1 \times 1$  kernel. Note that the number of channels is doubled after each pooling layer.

## 5.3 Fine-tuning details of proposed U-net architecture

### 5.3.1 Backbone U-Net architecture

The model architecture is an end to end deep learning approach that takes inspiration from the encoder, decoder backbone of U-Net, and the feature enricher Bidirectional feature network. The proposed model makes the use of U-Net based backbone network incorporated with a Bidirectional feature network for the task of uterus prediction. Further, the fully convolutional network-based U-Net encoder takes the MRI input images and outputs features at five corresponding depths. The feature network’s outputs are combined respectively with a decoder architecture to obtain a combination of lower-level fine-grained features with high-level semantic features [107]. The incorporation of a bidirectional feature network aims to improve the feature extraction efficiency at each level of the backbone architecture and enrich the feature vectors, thereby allowing a fusion of lower-level fine-grained features and higher-level semantic features.

### 5.3.2 Bidirectional Feature Network

The bidirectional feature network also incorporates additional weight for each input during feature fusion, thereby allowing the network to learn the particu-

lar input feature importance. For dynamic learning behavior and accuracy fast normalized fusion (one of the methods of incorporating weights during feature fusion) is implemented [108]. Also, for improvement of efficiency, depth-wise separable convolution followed by batch normalization and non-linear activation function ReLu (Rectified Linear unit) are implemented. This modification improves the feature maps at each depth of the network and provides an efficient fusion of features across various depths of the encoder section of the U-Net backbone architecture.

Instead of a single image, we input two images to the modified U-net: (i) the original pre-BT MRI in the first channel and (ii) the corresponding pre-BT uterus segmentation mask (binary image mask) in the second channel. U-net predicts the at-BT uterus segmentation mask (binary image mask). The uterine contour in the at-BT position is extracted from this binary mask. The complete U-net architecture is illustrated in Fig. 5.1.

The network training aims to increase probability of right class of each voxel in the mask. To accomplish this, a weighted binary cross-entropy loss of each sample of training has been utilized. For the implementation of weighted binary cross-entropy, the positive pixels by the ratio of negative to positive voxels in the training set was weighted. The network will learn to be less biased towards outputting negative voxels due to the class imbalance in the masks. The weighted binary cross-entropy loss is formulated as follows:

$$Loss = -\frac{1}{n} \sum_i [w_p \times y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (5.1)$$

where,  $n$  represents the number of samples,  $w_p$  represents the positive prediction weights and  $\hat{y}_i$  indicates the prediction of the model.

### 5.3.3 ReLU Activation Functions

In the neural network, the activation function is the gateway to incorporating non-linearity. It plays a pivotal part in the training and evaluation of deep neural networks. The widely used activation functions are ReLU, Sigmoid, Leaky ReLU, Tan hyperbolic. ReLU introduces the property of non-linearity to a deep learning model and solves the vanishing gradients issue. The main reason ReLU wasn't used until more recently is because it was not differentiable at the point zero. Researchers tended to use differentiable functions like sigmoid and tanh. However, it's now determined that ReLU is the best activation function for deep learning. The ReLU activation function is differentiable at all points except at zero. For values greater than zero, we just consider the max of the function. All the negative values default to zero, and the maximum for the positive number is taken into consideration. For the computation of the backpropagation of neural networks, the differentiation for the ReLU is relatively easy. The only assumption we will make is the derivative at the point zero, which will also be considered as zero. The main advantages of the ReLU activation function are:

- It is the most popular activation function for training convolutional layers and deep learning models.
- The rectifier function is trivial to implement, requiring only a max function.
- An important benefit of the rectifier function is that it is capable of outputting a true zero value.
- A neural network is easier to optimize when its behavior is linear or close to linear.

### 5.3.4 Data augmentation

Medical image segmentation is constrained by the abundant availability of labeled training data. Data augmentation helps to prevent the model from over-fitting and helps in improving the generalization capability of the network on data outside the training set. In medical imaging, the augmentations are provided to both the image and label equally, thereby creating warped versions of the training data. Common data augmentation techniques adopted in medical image analysis applications [89] include cropping, translation, rotation, flipping, and scaling of images. Here we rotated the MR images with different degrees (5, 10, 15, 20, 25, 30) clockwise and anti-clockwise to generate more training examples to train the DL models. At the same time, the same operation applied to an MR image is used for corresponding mask images. In this way, each augmented image pair’s relationship between the MR image and the output mask is preserved.

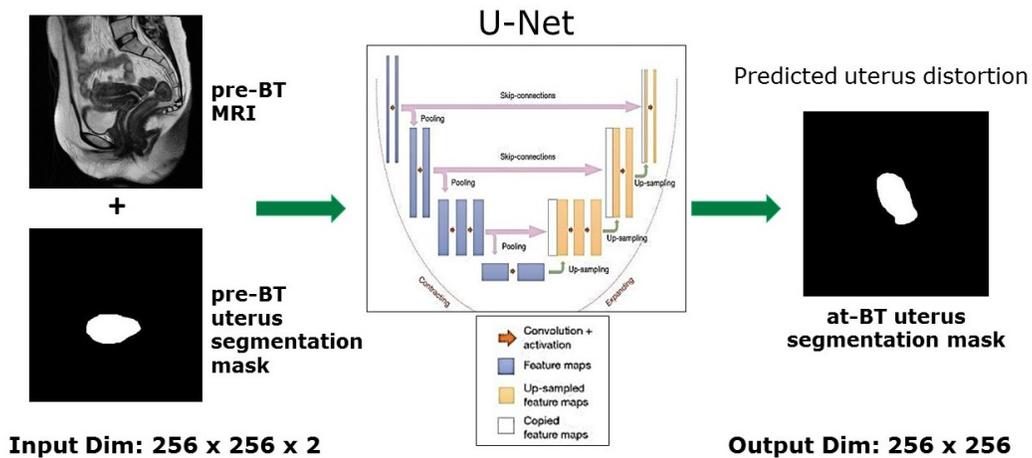


Figure 5.1: Architecture of the proposed U-net model for predicting the deformed at-BT uterus.

## 5.4 Comparison with U-Net++ Architecture

The state-of-the-art models for image segmentation are variants of the encoder-decoder architecture like U-Net and fully convolutional network (FCN). These encoder-decoder networks used for segmentation share a key similarity: skip connections, which combine deep, semantic, coarse-grained feature maps from the decoder sub-network with shallow, low-level, fine-grained feature maps from the encoder sub-network. To address the need for more accurate segmentation in medical images, another variation of U-Net architecture called U-Net++ is introduced in [109], a new segmentation architecture based on nested and dense skip connections. The underlying hypothesis behind this architecture is that the model can more effectively capture fine-grained details of the foreground objects. This is in contrast to the plain skip connections commonly used in U-Net, which directly fast-forward high-resolution feature maps from the encoder to the decoder network. U-Net++ starts with an encoder sub-network or backbone followed by a decoder sub-network. Each convolution layer is preceded by a concatenation layer that fuses the output from the previous convolution layer of the same dense block with the corresponding up-sampled output of the lower dense block. The main idea behind is to bridge the semantic gap between the feature maps of the encoder and decoder prior to fusion. The architecture of U-Net++ is illustrated in Figure 5.2.

U-Net++ takes advantage of re-designed skip pathways and deep supervision. The re-designed skip pathways aim at reducing the semantic gap between the feature maps of the encoder and decoder sub-networks, resulting in a possibly simpler optimization problem. With deep supervision: accurate mode wherein the outputs from all segmentation branches are averaged. U-Net++ generates full resolution feature maps at multiple semantic levels. The loss are estimated from 4 semantic levels. Also, a combination of binary cross-entropy

and dice coefficient as the loss function.

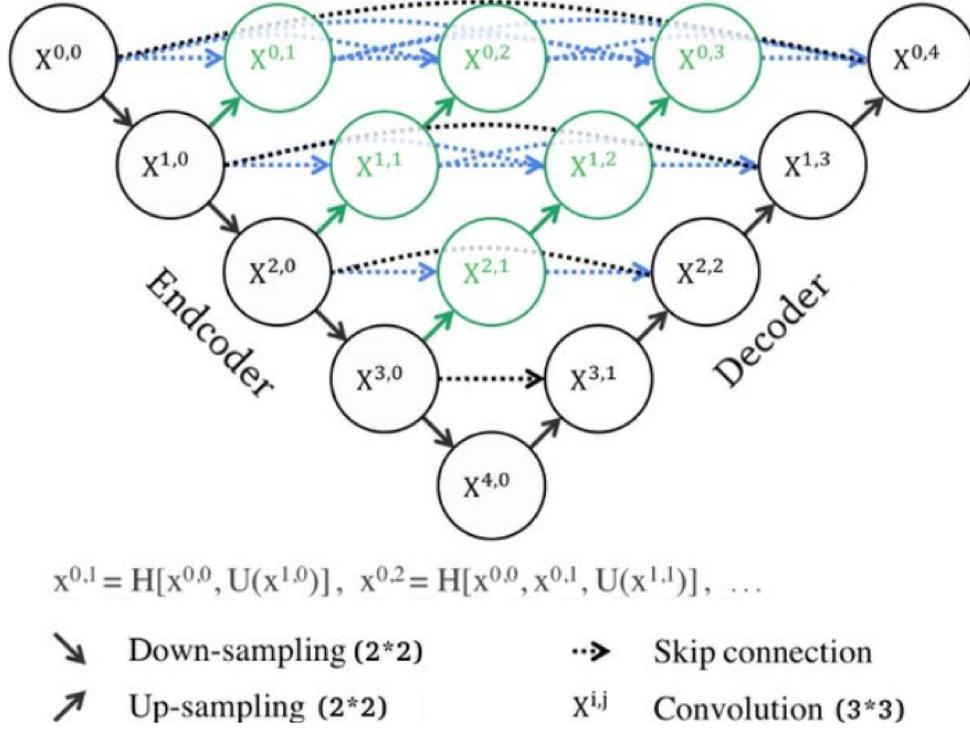


Figure 5.2: Illustration of U-Net++ Architecture

## 5.5 Results

### 5.5.1 Implementation Details

The training approach utilizes 5-fold cross-validation to obtain an accurate measure of the generalizing capability of the proposed model [110]. The Adam version of stochastic gradient descent (SGD) trained the proposed network. The regularization parameter was  $\lambda = 10^4$ . The learning rate was set from  $10^4$  to  $10^5$ , which was gradually reduced at each epoch. The number of the epoch was 250. In this experiment, the ReLU activation function has been used for efficient training of the model, and also implementation of data augmentation was done on the training set to improve the robustness of the model.

The image size was  $256 \times 256$ . All the experiments with the neural network were implemented using Keras API [95] with TensorFlow and on two NVIDIA GeForce GTX 1080 GPU processors. It took around 5 hours to train the network. The prediction took 4 ms per image.

## 5.5.2 Automated Uterus Structure Prediction

### Using CNN

Table 5.1: Evaluation of uterus prediction results by the proposed method in terms of Dice Coefficient (DC) and Hausdorff Distance (HD) in comparison to expert manual segmentation. The Mean  $\pm$  Standard Deviation values are reported for each metric. The higher the Dice coefficient or the lower the Hausdorff distance the better the results.

	Dice Coefficient (%)	Hausdorff Distance (mm)
<b>U-Net</b>	<b><math>89.5 \pm 4.2</math></b>	<b><math>3.6 \pm 2.6</math></b>
<b>U-Net++</b>	$87.8 \pm 5.3$	$6.2 \pm 3.7$

The conventional U-net model predicts the at-BT uterus structure from pre-BT MR image and pre-BT uterus segmentation. The proposed approach yielded an average Dice score of 89.5% and an average Hausdorff distance of 3.6 mm compared to the ground truth segmentation. The results of U-Net is compared with U-Net++ and is shown in Table 5.1.

The MSE loss curves (training and validation curves) with different epochs are shown in Figure 5.3.

In Figure 4.8, the highest accuracy (90%), 50th percentile (87%), 25th percentile (83%) and the lowest accuracy (78%) of Dice scores achieved by the proposed method are shown visually, where the green, yellow and red contours depict the segmentation results by the proposed method (CNN with autoencoder), U-net approach and manual ground truth, respectively.

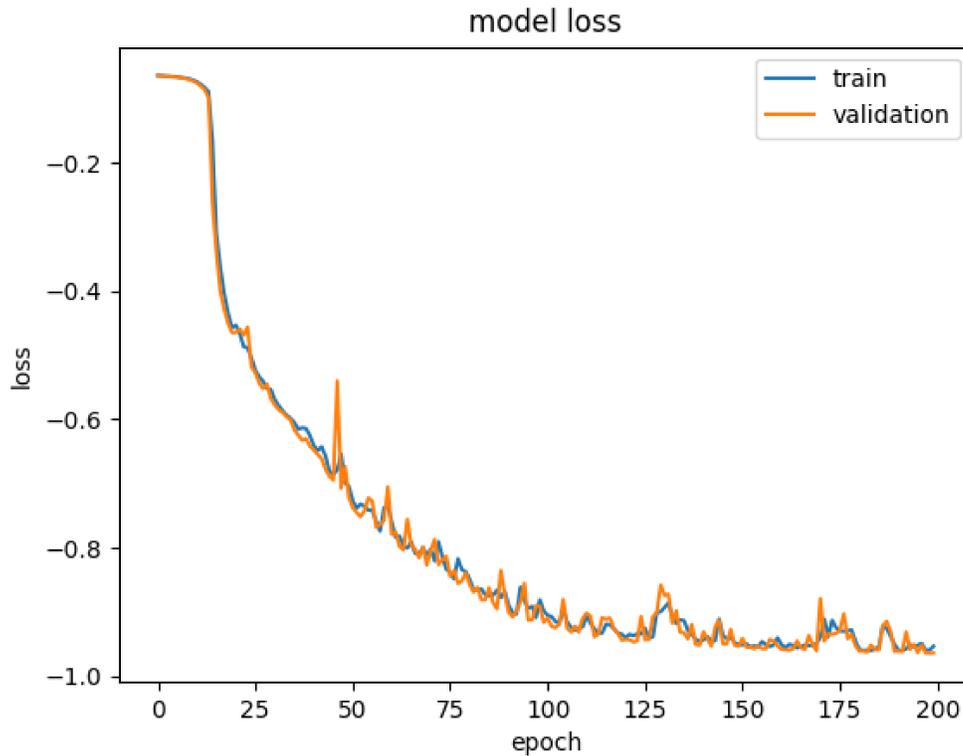


Figure 5.3: Loss - (MSE) vs. Epoch curve for training and validation for U-net

## 5.6 Discussion

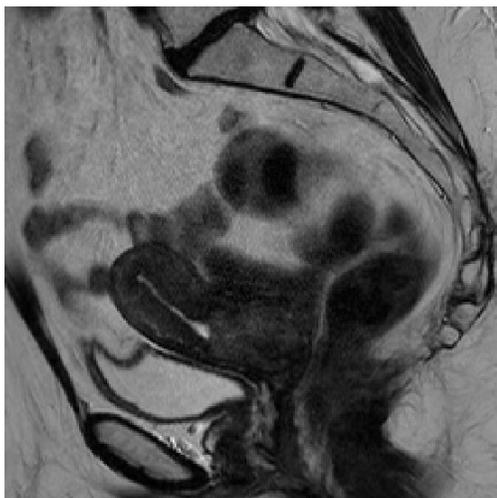
The proposed DL-based framework can be incorporated as an automatic prediction tool of uterine deformation due to applicator insertion for personalized BT treatments for cervical cancer patients. It holds promise for more streamlined clinical/technical decision-making before BT applicator insertion resulting in improved dosimetric outcomes. Before the BT procedure, the physicians make several educated predictions, based on the pre-BT MRIs, to select the appropriate applicator for each patient. However, with large and complex distortions caused by applicator insertion, such speculation to guide implant strategy is far from accurate, affecting implant quality and treatment outcomes. The problem that we have addressed in this thesis, i.e. predicting the anatomy deformation in cervical cancer brachytherapy caused by inserting



(a) preBT MRI



(b) atBT Uterus prediction result: DC 0.90



(c) preBT MRI



(d) atBT Uterus prediction result: DC 0.80

Figure 5.4: Results with different Dice score values for predicting the uterus structure using U-net (green) and U-Net++ (yellow). The red contour represents the manual ground truth. The corresponding input MRI (pre-BT) is also displayed in the right to have a clear idea about the problem formulation and the output results.

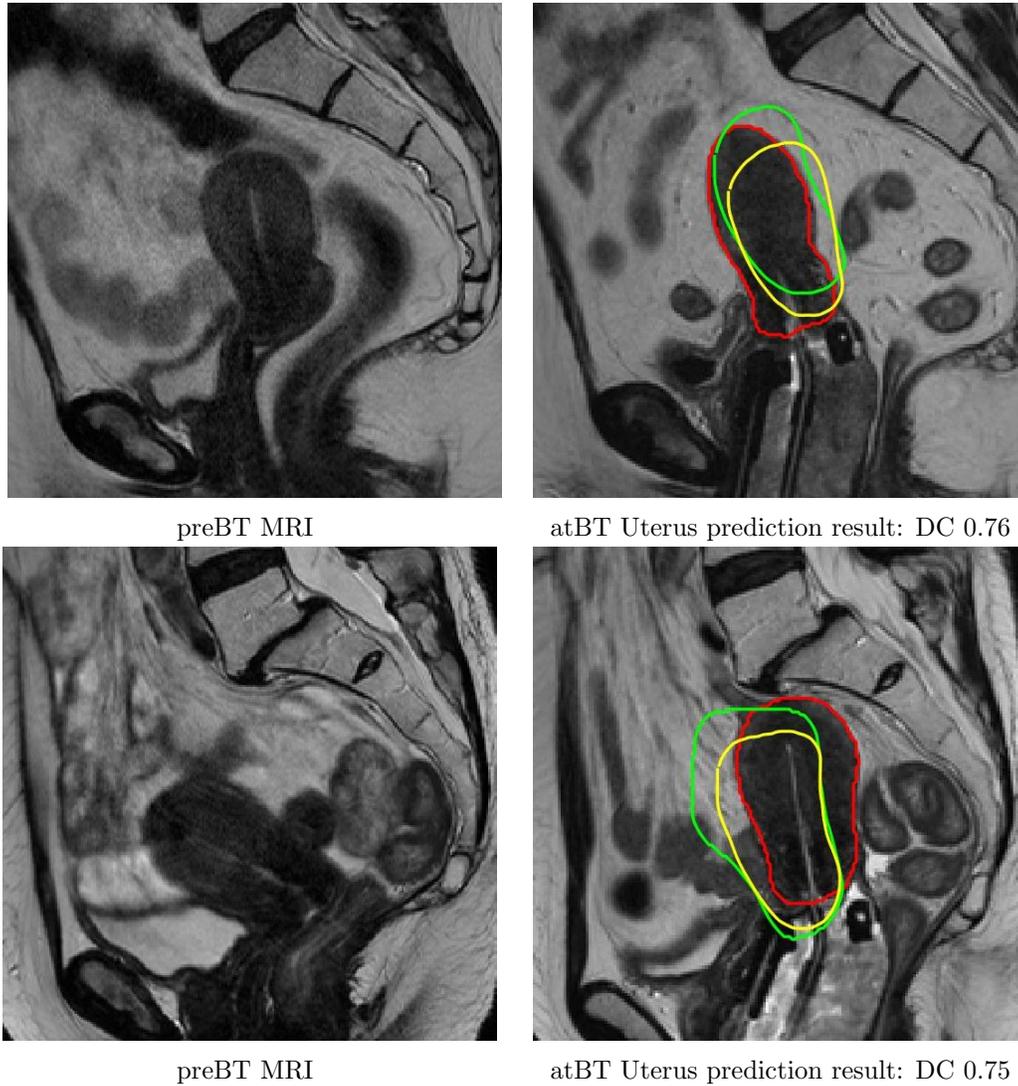
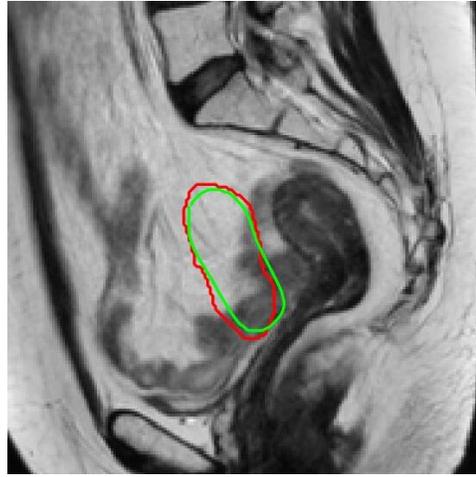
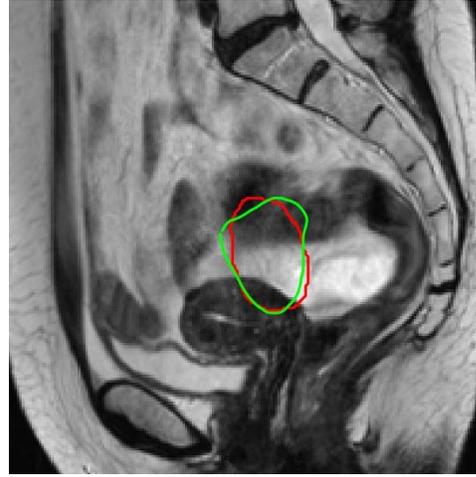


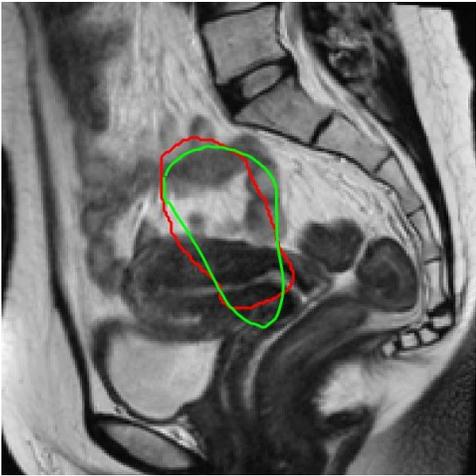
Figure 5.5: Results with different Dice score values for predicting the uterus structure using U-net (green) and U-Net++ (yellow). The red contour represents the manual ground truth. The corresponding input MRI (pre-BT) is also displayed in the right to have a clear idea about the problem formulation and the output results.



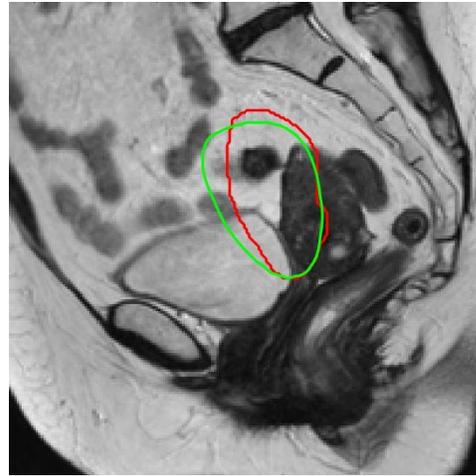
(a) Highest DC - 0.90



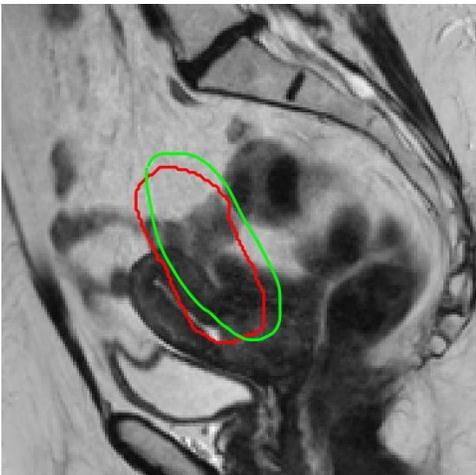
(b) DC - 0.88



(a) 50th percentile - 0.86



(b) 25th percentile - 0.82



(c) DC - 0.80



(d) Lowest dice score - 0.78

Figure 5.6: Results with different Dice score values for predicting the uterus structure using U-net (green) and manual ground truth (red). As for any test cases, there is no at-BT image, that is why

the BT applicator, is unique and has not yet been reported by any other study. Our proposed method (modified U-Net) achieved 89.5% accuracy in terms of Dice Coefficient (DC) in predicting uterine shape deformation and outperformed U-Net++, which is quite acceptable as there is no other method available now in medical imaging that can solve this problem. Where as U-Net++ achieved 87.8% DC which is comparable with our proposed method, but for a few difficult cases our approach with modified U-Net performed significantly better than U-Net++. Figure 5.4 and Figure 5.5, display the visual results of U-Net and U-Net++. From the results we can conclude that, although the measures are comparable but in most of the cases U-Net outperformed U-Net++ in predicting at-BT uterus structure from pre-BT MRI.

# Chapter 6

## Non-rigid registration/Free form deformation

### 6.1 Overview

Medical image registration is one of the most important and challenging research in the modern medical image analysis field. From the view of the image transformation, medical image registration can be classified into rigid registration and non-rigid registration. As our goal is to predict the uterine deformation in cervical cancer brachytherapy, that is why we are focusing on non-rigid registration method using free form deformation. The non-rigid medical image registration algorithm naturally depends on the geometric deformation model and the similarity measure criterion. The geometric deformation models can be classified into two main categories: i) physics-based models such as the elastic body models [111,112], the optical flow models [113] and the diffusion models [83]; and ii) interpolation-based models such as free-form deformations [114].

For the above registration algorithms based on the physical models, it is

difficult to construct a reasonable physical model that can simulate the complex tissue deformations between the two input images. In [114], the authors proposed a local deformation model for non-rigid registration on breast MR images. This model was described by the so called Free-Form Deformation (FFD) based on B-splines, and it employed the normalized mutual information (NMI) as the similarity function. Since the degree of freedom of the local deformation model is determined by the number of control points, it is important to decide whether a sparse or dense set of control points should be used. However, both sparse and dense sets have limitations. If a sparse set of control points is used, the movements of the control points will not well represent complicated deformations. If a dense set of control points is used, the optimization can be computationally inefficient. In order to tackle these shortcomings, some researchers proposed some alternate methods. For example, [115] proposed a multi-level B-spline model in which only a sparse subset of the control points is active to balance speed and accuracy.

Free Form Deformation (FFD) is part of the computer graphics literature on soft objects. The definition of a soft object is an object that can be deformed by the user or during the process of animation. Soft object deformation is used for many purposes: 1) Shape distortion to highlight dynamic interaction with the environment. For instance, an animator may want to create a basketball that will deform when it bounces on the ground. Another use would be to deform the shape of a car during a collision in a racing simulation. 2) Realistic deformation of an object that has a highly elastic and flexible shape. Examples include the facial expressions, motion of the human body, and cartoon animation etc.

The FFD model employs a B-spline mesh to estimate the deformation field. The mesh is controlled by the grid points, referred to as control points, which

are scattered in a regular spacing grid. Each control point is associated with a kernel function, which defines the local deformations induced by the displacement of this control point. A popular kernel function chosen is the cubic B-spline kernel function. Here, after predicting the at-BT uterus shape and location for the pre-BT MRI image using deep convolutional neural networks, we need to determine the amount of deformation or shift that happened in the uterus structure due to the applicator insertion. This information will be helpful for the clinicians who perform the brachytherapy treatment and will guide them to understand the anatomy deformation in a better way. For example, suppose the clinicians have the information about the possible amount of uterine deformation before inserting the applicator. In that case, this will help them select the correct applicator configuration for each patient. To measure the amount of deformation, we used a non-rigid shape-based registration method and a free-form deformation approach. The detailed methodology is explained in the next section below.

## 6.2 Methodology

The shape registration problem has been studied mainly in the literature and represents a fundamental problem in different computer vision and image processing applications. It tries to recover a set of transformation parameters that brings a given **data set** as close as possible to the corresponding **model set**. The rigid case, also known as shape alignment, involves simple rotations and translations. While in the non-rigid shape registration case, in addition to the rotation and translation parameters, it includes a set of deformation parameters. In general, most of the approaches proposed for non-rigid shape registration follow a two-step scheme, where first, a rigid global alignment is

performed. Then a local process deforms the shape of the data set towards the given model set.

We tested a shape registration algorithm proposed by [116] that tackles global and local alignment and deformation. The method follows these steps: (1) a robust distance approximation based on local curvature information is used for non-rigid registration; (2) the proposed objective function is in the linear least squares form, hence it can be solved by a linear system of equations; (3) the proposed method captures all deformation from rigid to non-rigid by the same framework; there is no need to use different steps to capture global and local deformations separately; (4) unlike the sign distance field, the proposed function is not discretely approximated.

### 6.2.1 Registration Error (Squared Distance)

Signed distance fields (SDFs) are used in [117] to capture the local transformation in small sampling grids. The FFD control lattice defines these sampling grids. The approximation error estimates the distance between the current data set and the model set. All the registration methods seek the best transformation parameters to move the given *data set*  $S = \{s_i\}$  (source shape) close to the *model set*  $T = \{t_j\}$  (target shape).

A well-known example method is the Iterative Closest Point (ICP) algorithm to measure the distance between the data set and model set. It moves the data set in each iteration based on a simple criterion: for the given data point, it searches for the closest corresponding model point. Therefore, the distance used by ICP is a point-to-point distance, and ICP performs a Point Distance Minimization (PDM) in each iteration to find the best transformation parameters [118]. One scenario could be when some of the data points lie on the curve passing through the model set; hence their distance to the model

set must be quite low, but ICP devotes a quite high distance to these points since the model set is quite sparse. If there could be a better approximation for the distance, the ICP would devote more weight to the data point, which is still far from the model point.

Here we used a quadratic approximation of the geometric distance to define the registration error term in the least squares form. This distance is based on the curvature information in the model sets. Consider the data point  $s_i$  with its closest corresponding model point  $t_j$ . Then the Squared Distance (SD) of  $s_i$  to the whole model set  $T$  can be approximated as follows:

$$SD(s_i, T) = \frac{d}{(d - \rho)} [(s_i - t_j) \cdot T_j]^2 + [(s_i - t_j) \cdot N_j]^2 \quad (6.1)$$

where  $T_j$  and  $N_j$  are the unit tangent and unit outer normal, respectively, defined in the Frenet–Serret frame at  $t_j$ . The value  $\rho$  is the curvature radius at the model point  $t_j$ , and  $d$  is the signed distance between the data point  $s_i$  and the model point  $t_j$ . The sign of  $d$  is positive if  $s_i$  and  $N_j$  lie on the same side and are negative otherwise.

The squared distance approximation works with the Frenet–Serret frame at the foot-point  $t_j$ . It projects the data point on the normal and tangent vectors first, and the final approximation will be quadratic with respect to these projections. In the particular case where the data point is along the normal at the foot-point, the first quadratic term vanishes and the distance will be equal to  $|s_i - t_j|^2$ , which is the squared point-to-point distance. In another case, where the model’s curvature set at  $t_j$  is zero, the first quadratic term vanishes again, and the projection of the data point on the normal will be the SD approximation of quadratic distance.

## 6.2.2 Free Form Deformation

The shape registration problem has been studied mainly in the literature and represents a fundamental problem in different computer vision and image processing applications. It tries to recover a set of transformation parameters that brings a given **data set** as close as possible to the corresponding **model set**. The rigid case, also known as shape alignment, involves simple rotations and translations. While in the non-rigid shape registration case, in addition to the rotation and translation parameters, includes a set of deformation parameters. In general, most of the approaches proposed for non-rigid shape registration follow a two-step scheme, where first, a rigid global alignment is performed. Then a local process deforms the shape of the data set towards the given model set.

We tested a shape registration algorithm proposed by [116] that tackles global and local alignment and deformation. The method follows these steps: (1) a robust distance approximation based on local curvature information is used for non-rigid registration; (2) the proposed objective function is in the linear least squares form, hence it can be solved by a linear system of equations; (3) the proposed method captures all deformation from rigid to non-rigid by the same framework; there is no need to use different steps to capture global and local deformations separately; (4) unlike the sign distance field, the proposed function is not discretely approximated.

## 6.2.3 Registration Error (Squared Distance)

Signed distance fields (SDFs) are used in [117] to capture the local transformation in small sampling grids. The FFD control lattice defines these sampling grids. The approximation error estimates the distance between the current data set and the model set. All the registration methods seek the best trans-

formation parameters to move the given *data set*  $S = \{s_i\}$  (source shape) close to the *model set*  $T = \{t_j\}$  (target shape).

A well-known example method is the Iterative Closest Point (ICP) algorithm to measure the distance between the data set and model set. It moves the data set in each iteration based on a simple criterion: for the given data point, it searches for the closest corresponding model point. Therefore, the distance used by ICP is a point-to-point distance, and ICP performs a Point Distance Minimization (PDM) in each iteration to find the best transformation parameters [118]. One scenario could be when some of the data points lie on the curve passing through the model set; hence their distance to the model set must be pretty low, but ICP devotes a high distance to these points since the model set is quite sparse. If there could be a better approximation for the distance, the ICP would save more weight on the data point, which is still far from the model point.

Here we used a quadratic approximation of the geometric distance to define the registration error term in the least squares form. This distance is based on the curvature information in the model sets. Consider the data point  $s_i$  with its closest corresponding model point  $t_j$ . Then the Squared Distance (SD) of  $s_i$  to the whole model set  $T$  can be approximated as follows:

$$SD(s_i, T) = \frac{d}{(d - \rho)} [(s_i - t_j) \cdot T_j]^2 + [(s_i - t_j) \cdot N_j]^2 \quad (6.2)$$

where  $T_j$  and  $N_j$  are the unit tangent and unit outer normal, respectively, defined in the Frenet–Serret frame at  $t_j$ . The value  $\rho$  is the curvature radius at the model point  $t_j$ , and  $d$  is the signed distance between the data point  $s_i$  and the model point  $t_j$ . The sign of  $d$  is positive if  $s_i$  and  $N_j$  lie on the same side and are negative otherwise.

The squared distance approximation works with the Frenet–Serret frame

at the foot-point  $t_j$ . It projects the data point on the normal and tangent vectors first, and the final approximation will be quadratic with respect to these projections. In the particular case where the data point is along the normal at the foot-point, the first quadratic term vanishes and the distance will be equal to  $|s_i - t_j|^2$ , which is the squared point-to-point distance. In another case, where the model’s curvature set at  $t_j$  is zero, the first quadratic term vanishes again, and the projection of the data point on the normal will be the SD approximation of quadratic distance.

### 6.3 Results

The performance of the shape deformation approach has been evaluated on a dataset of 120 cervical cancer patients. Here the **data set** corresponds to the pre-BT shape, and the **model set** conforms to the at-BT shape. The uterine canal, vaginal canal and uterus shapes have been deformed using this approach.

Table 6.1: Quantitative analysis of anatomical deformation (uterus, uterine canal, vaginal canal) due to applicator insertion between pre-BT and at-BT anatomical structures. The Mean  $\pm$  Standard Deviation values of the Mean Squared Distance are reported here.

<b>Anatomy deformation estimation (mm)</b>	
<b>Uterus</b>	$35.8 \pm 5.2$
<b>Uterine Canal</b>	$30.7 \pm 6.9$
<b>Vaginal Canal</b>	$28.4 \pm 4.6$

The following figures show illustrations of 2D shapes (uterine canal and vaginal canal, and uterus) registered with the proposed approach. In the current implementation, the regularization parameter  $\lambda$ , which somehow represents the registration rigidity, was automatically tuned. It starts with a high

regularization value  $\lambda = 10^5$ , mainly devoted to tackling the alignment problem. The figures depict intermediate results and the results after convergence are reached.

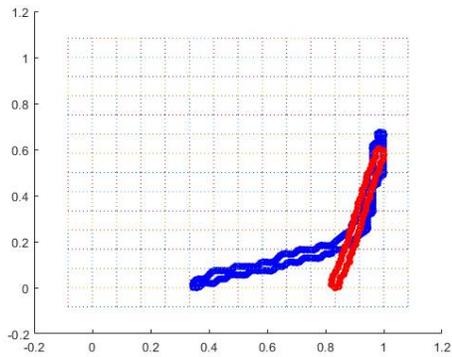
Figure 6.1 shows illustrations of a 2D uterine canal and the applicator model registered with the proposed approach. Figure 6.2 illustrates the 2D vaginal canal deformed according to the applicator. Figure 6.3 and Figure 6.4 show the deformation of the 2D uterus structure from the pre-BT position to the at-BT position.

## 6.4 Discussion

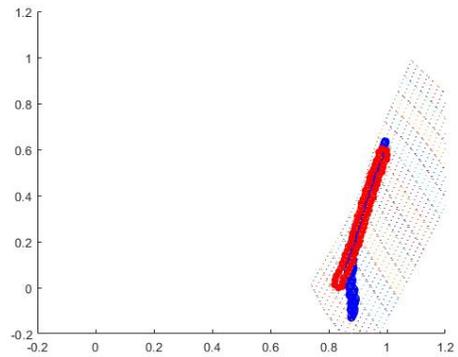
The tandem applicator is inserted through the patient's vaginal canal into the uterine canal. As we have the 2D shape of the applicator from the beginning as the reference shape, we tried to deform the 2D shape of the uterine canal and vaginal canal according to the applicator's shape (shown in Figure 6.1 and Figure 6.2). During this deformation procedure, we measured the amount of deformation that happened to the shapes of the canals. The applicator insertion induced uterine and vaginal canal displacement of 38.6 mm and 30.5 mm, respectively.

In Figure 6.3 and Figure 6.4, the uterus deformation from pre-BT to at-BT treatment for two different patients is demonstrated. Large displacements and deformations were observed in these cases.

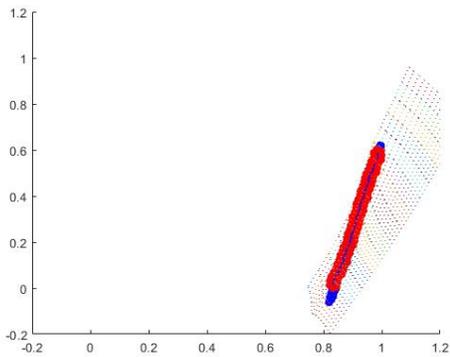
The applicator insertion induced the largest mean displacement for the uterus of 29.5 mm and 62.3 mm, respectively, for these two patients. Shape with red depicts the pre-BT anatomy, and the yellow shape represents the at-BT anatomy. The red shape is deformed to the yellow shape using the free-form deformation technique.



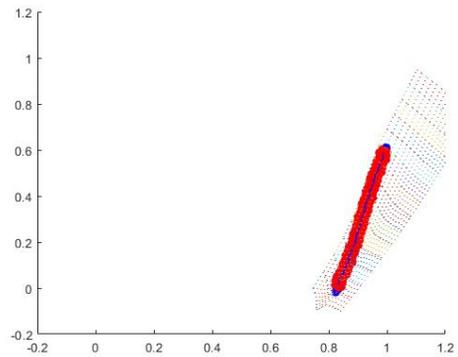
(a) Iteration 1 : 29.8 mm



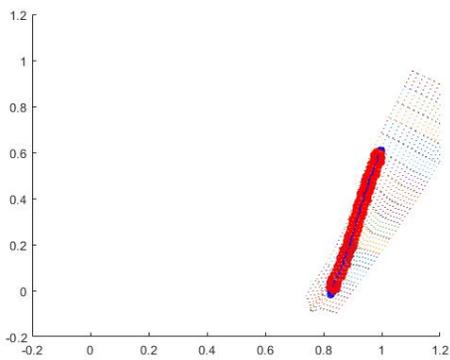
(b) Iteration 2 : 6.4 mm



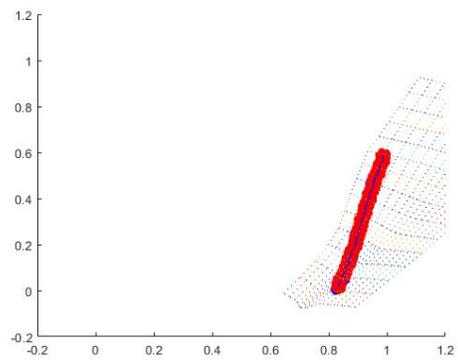
(c) Iteration 3 : 3.1 mm



(d) Iteration 4 : 2.4 mm

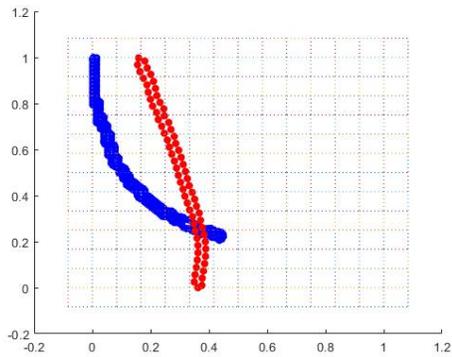


(e) Iteration 5 : 1.7 mm

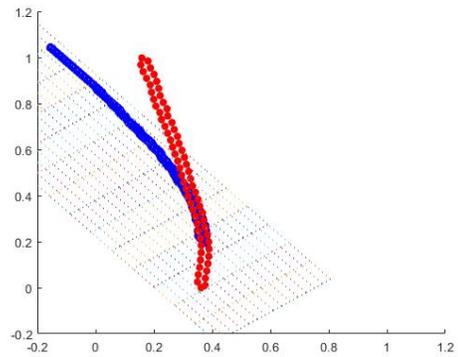


(f) Iteration 6 : 1.3 mm

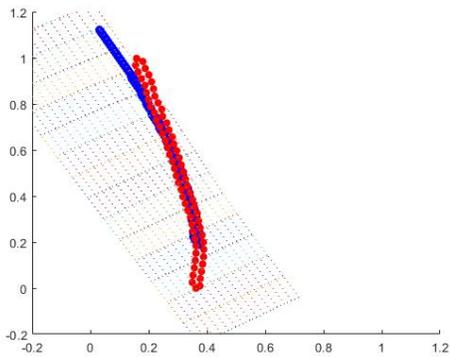
Figure 6.1: Registering the **pre-BT uterine canal** with the applicator shape. Red - applicator shape and Blue - pre-BT uterine canal. The iteration number and the distance between pre-BT and at-BT shapes are given here.



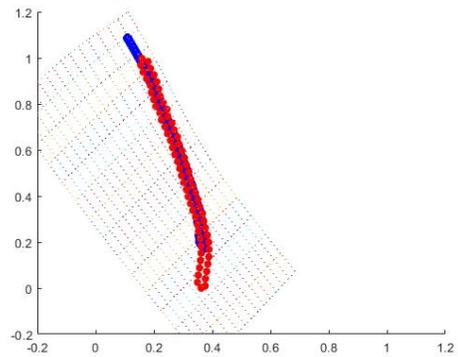
(a) Iteration 1 - 29.8 mm



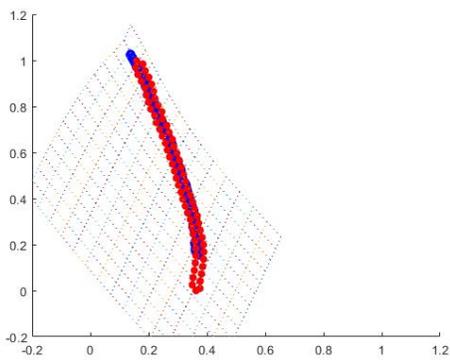
(b) Iteration 2 - 18.7 mm



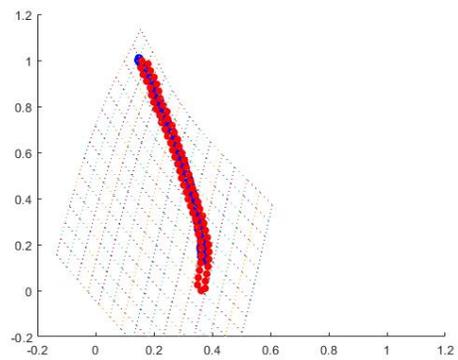
(c) Iteration 3 - 6.1 mm



(d) Iteration 4 - 3.5 mm

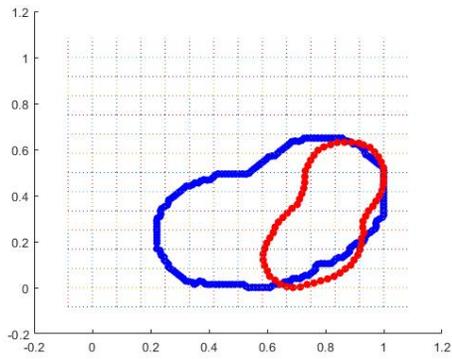


(e) Iteration 5 - 2.1 mm

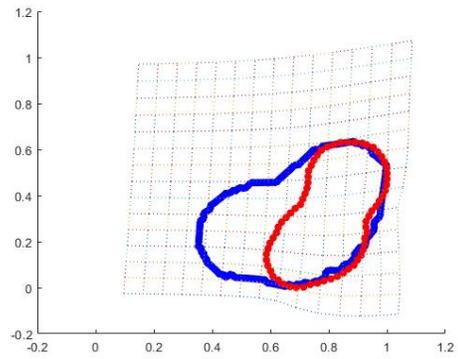


(f) Iteration 6 - 1.7 mm

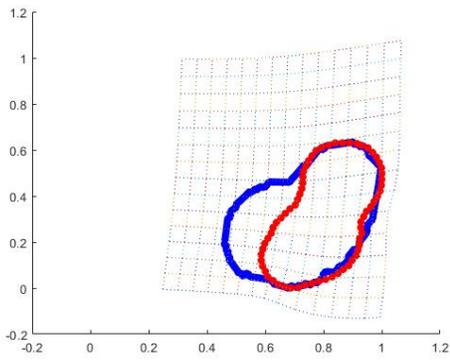
Figure 6.2: Registering the **pre-BT vaginal canal** with the applicator shape. Red - applicator shape and Blue - pre-BT vaginal canal. The iteration number and the distance between pre-BT and at-BT shapes are given here.



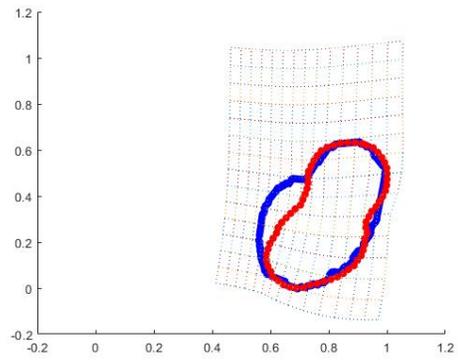
(a) Iteration 1 - 29.4 mm



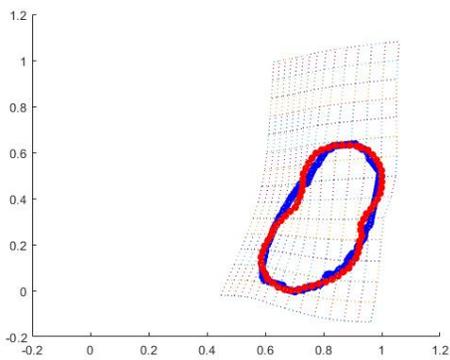
(b) Iteration 2 - 17.2 mm



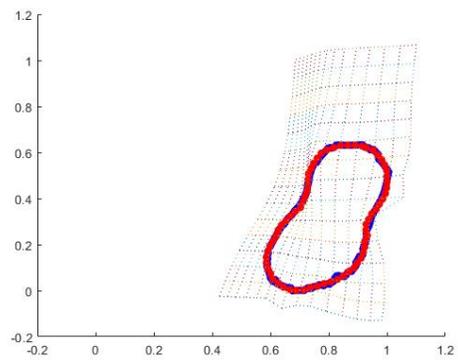
(c) Iteration 3 - 9.9 mm



(d) Iteration 4 - 5.1 mm

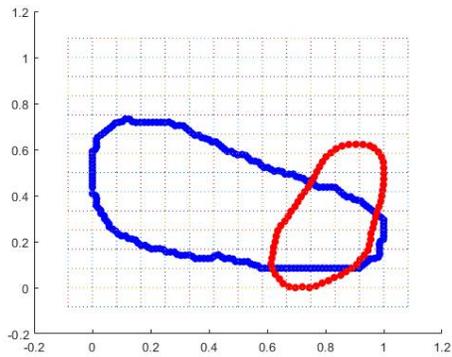


(e) Iteration 5 - 2.6 mm

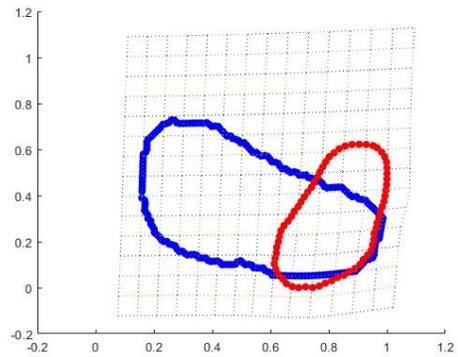


(f) Iteration 6 - 1.2 mm

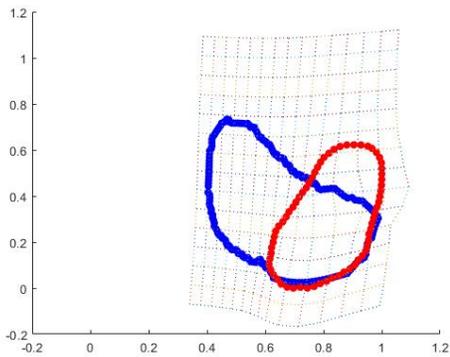
Figure 6.3: Registering the **pre-BT uterus** with at-BT uterus. Red - at-BT uterus and Blue - pre-BT uterus. The iteration number and the distance between pre-BT and at-BT uterus shapes are given here.



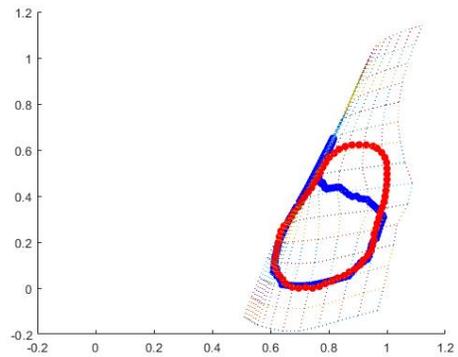
(a) Iteration 1 - 62.5 mm



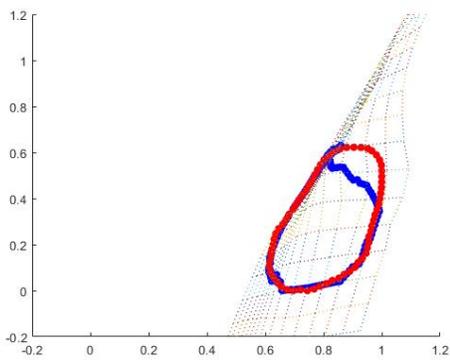
(b) Iteration 2 - 48.2 mm



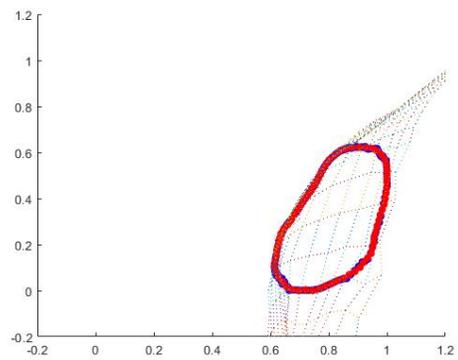
(c) Iteration 3 - 29.6 mm



(d) Iteration 4 - 13.5 mm



(e) Iteration 5 - 4.1 mm



(f) Iteration 6 - 2.6 mm

Figure 6.4: Registering the **pre-BT uterus** with at-BT uterus. Red - at-BT uterus and Blue - pre-BT uterus. The iteration number and the distance between pre-BT and at-BT uterus shapes are given here.

Predicting and measuring deformation from the natural anatomy prior to BT to anatomy in the presence of an intrauterine applicator is challenging as the uterus, and surrounding organs, deform in a unique way relative to each other, and for every patient. However, the applicator model and its geometry are fixed usually. In clinical practice, radiation oncologists make several experience-informed predictions prior to BT applicator selection and insertion, partly based on pre-BT MRI, gauging potential uterine deformation, the final decision of the chosen tandem relative to tumour targets, size of the intravaginal applicator, and radial and deep positions for applicators. An implant strategy is formulated for anatomic and geometric fit, which also considers anticipated dosimetry. The possibility to correctly predict uterine distortion using the DL-method demonstrates a first milestone towards achieving these technical BT parameters with improved accuracy and ease. The challenge in implementing this algorithm was the inter-patient anatomical dissimilarity and extreme intra-patient uterine deformation from pre-BT to at-BT in the dataset. Increasing the size of our training dataset, with the inclusion of more heterogeneous images with anatomical variability, will improve the prediction accuracy of this DLbased algorithm. These uterine deformation measurements will guide the clinicians in accurate prediction of the applicator geometry for each LACC patient.

# Chapter 7

## Conclusion and Future Work

### 7.1 Discussion

#### 7.1.1 Application of Mass-Spring Methods (MSMs)

The problem of predicting anatomical deformation due to applicator insertion is a very rare and difficult problem in the medical imaging field. The ideal solution needs simulations of soft tissues and internal organs in a proper scientific and medical environment. Various simulation methods and improvement approaches for modeling deformable bodies were explored [74, 119]. The methods aim for fast and robust simulations with physically accurate results. But most of the modeling approaches demand 3D data to perform the simulation of the internal organs like uterus, vagina, cervix etc. The main limitation of our dataset is that, we only have 2D MRI images of pelvic region (pre-BT and at-BT MRIs) to predict the uterine deformation due to applicator insertion in LACC BT. It is very challenging to develop a simulation model of the internal organs of pelvic region for one patient only from 2D MRI slice. Even if one can contour all the anatomical structures on the 2D MRI, the main concern will be to derive a good 3D view from it.

To develop a physical model, based on elasticity property of the uterus to deform it from its natural anatomy, a mass-spring damper system using the common material properties such as Young's Modulus and Poisson's Ratio were applied. A mass-spring model based shape deformation algorithm was investigated to deform the uterus structure from pre-BT to at-BT position. By defining a mesh model and adding a new generalized spring for each mass, our surface mesh model can preserve its original geometric features such as volume and shape. Global deformations such as stretch and shear, as well as rigid motions are separated from the physically based mass-spring model. They are modeled graphically and this approach largely improves the global deformation effects of the model.

In this work, the mass-spring model was utilized using Statistical and Machine Learning Toolbox of Matlab. The real-time computational complexity of our model is linear with the number of point masses, which is much more efficient than general FEMs. Furthermore, our model almost does not need any pre-processing because mass-spring models do not need the computation for global stiffness matrix which is needed in FEMs. The results are shown in Figure 7.1 and Figure 7.2.

The elastic properties of the organs or soft tissues are defined by two parameters (elastic moduli). Classically Young's modulus  $E$  and Poisson's ratio  $\nu$  are the popular pair. The Young's modulus is the ratio of stress to strain measured along the same axis under an uni axial stress condition, that is, it gives the resistance to directional stretching or compression. The Poisson's ratio is the ratio of transverse to axial strain (denotes to what degree material expands in one direction when compressed in another). Depending on the application, besides  $E$  and  $\nu$ , other moduli are often used such as bulk modulus  $K$ , or Lamé parameters  $\lambda$  and  $\mu$  [120]. In any description only two of them

are independent and providing a link between spring-network parameters, and a chosen pair of the elastic moduli is sufficient to describe elastic properties of the MSM. In case of 2D MSM, an isotropic homogeneous structure can be obtained with hexagonal lattice [120]. All the springs have the same spring coefficient  $k$  and the relation between the spring coefficient and the Lamé constants for such network is given by

$$\lambda = \mu = \frac{3}{4\sqrt{3}}k \quad (7.1)$$

from which it follows that  $E = \frac{2}{\sqrt{3}}k$  and  $\nu = \frac{1}{3}$ . Springs are assumed to be of a unit length.

### **Limitations of MSM in our dataset**

A stable and accurate deformable model to simulate the dynamics of soft tissues is a challenging area of research. Real-time and precise simulation of soft tissue deformation is still a major challenge. A realistic estimation of stiffness parameters (Young's modulus, shear modulus, bulk modulus, viscosity) is required as basis for the simulation models. The reliable approximation of these input parameters is difficult. FEMs and MSMs are most suitable when we have the appropriate data and the accurate values of the parameters which are responsible for the anatomy deformation.

As our original problem is predicting uterine deformation, there are various factors that affect soft tissue deformation. Elasticity property of the uterus, surrounding fluid density, presence of other organs affect the deformation immensely from patient to patient. We have seen from our data that, every patient's anatomy is different and it deforms in a distinctive way. So, predicting this deformation before applicator insertion is very much challenging even for the expert clinicians. Hence, we wanted to utilize deep learning methods

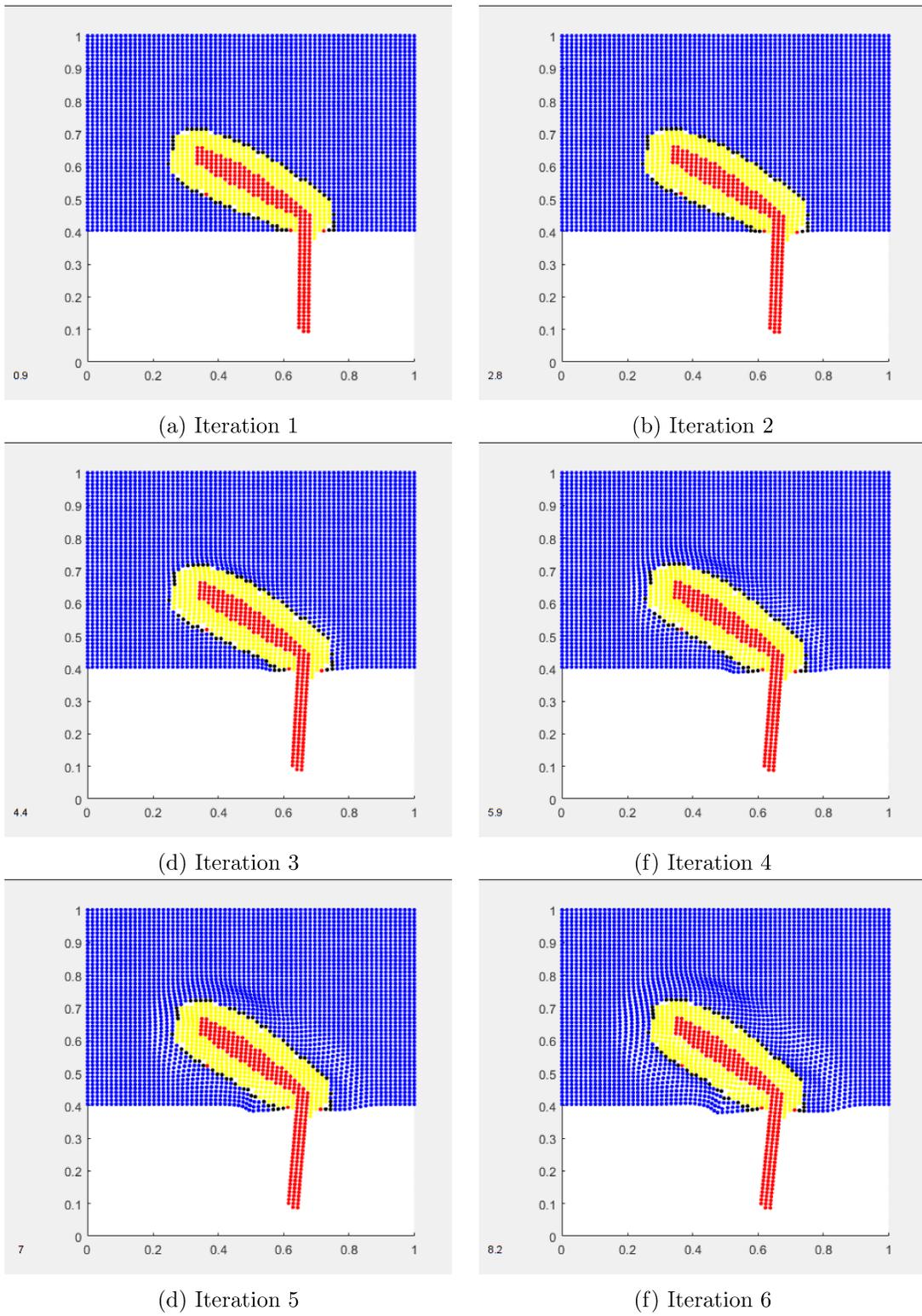


Figure 7.1: Results of uterus deformation using Mass-Spring Model where yellow structure is the uterus and red structure is the applicator model

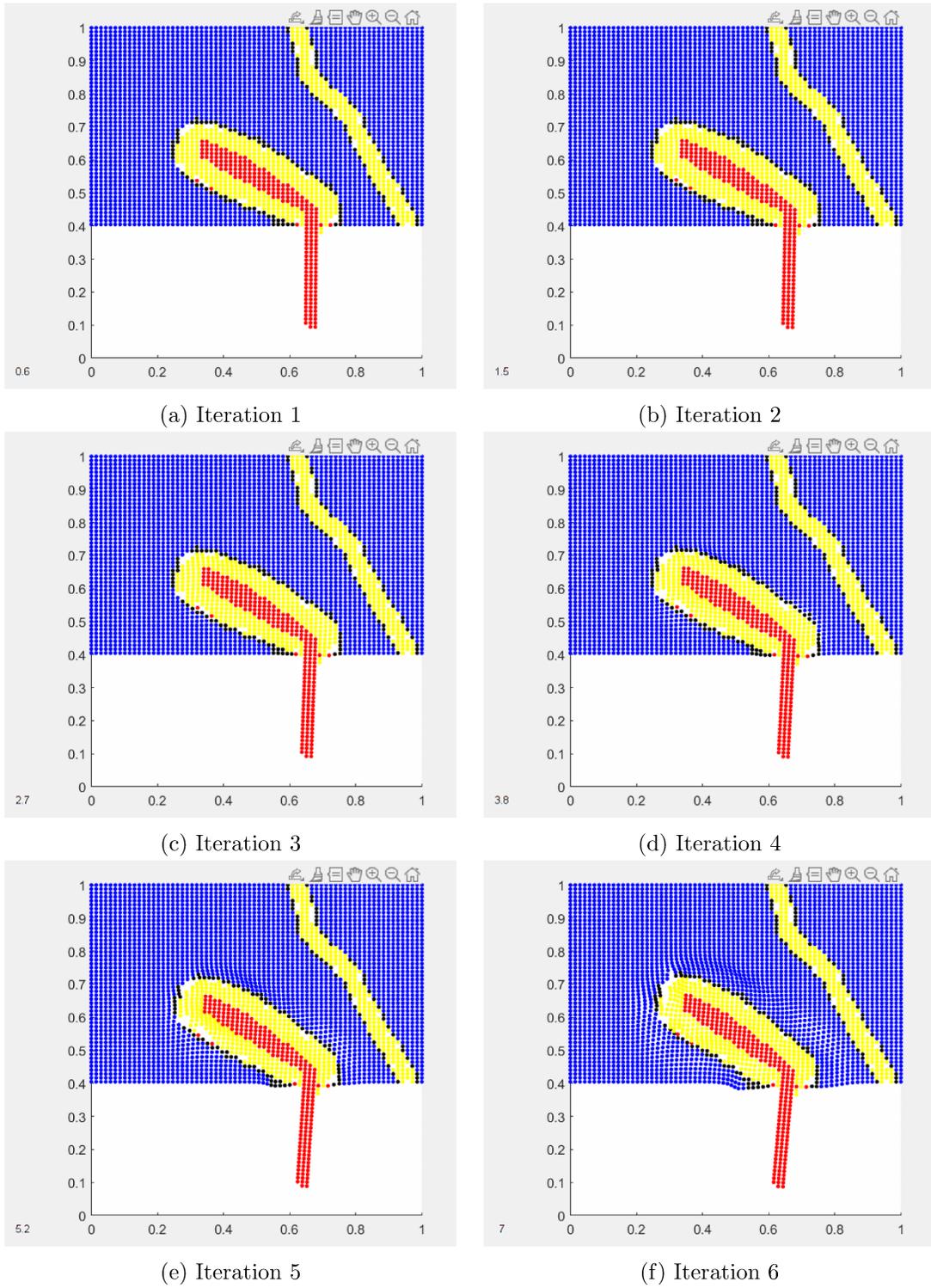


Figure 7.2: Results of uterus deformation using Mass-Spring Model in presence of the rigid bone structure

and also evaluate whether DL networks can discover the hidden information from the MRI data with which we can come up with a method to predict the anatomy deformation in cervical cancer brachytherapy.

### **Biomechanical properties of Uterus, Vagina, Cervix**

Soft tissues connect and support other tissues and surround the organs in the body. They include muscles (including the heart), fat, blood vessels, nerves, tendons, and tissues that surround the bones and joints. Soft tissues can be categorized into the skin, muscles, connective tissues, and various functional organs. The major soft tissues existing across the length and breadth of the human body. Due to trauma, cancer, or pathological conditions, soft tissues such as skin, vascular tissues, neural tissues, skeletal muscle, cartilage, and ligament become deficient or lose their natural functionality. In our case, the organs of interest are uterus, vagina, cervix for LACC BT treatment.

The uterus, through the cervix, is connected to the vagina and the fallopian tubes. This tissue is usually pear-shaped and is about 7.6 cm long, 4.5 cm wide, and 3.0 cm thick [121]. It undergoes large deformation during brachytherapy applicator insertion. Dias et al in [122] stated the elastic modulus and Poisson's ratio of the uterus tissue as 5 and 0.49 kPa, respectively. Baah-Dwomoh et al in [123] estimated the elastic modulus of the uterus and the associated ligaments. The reported values for the elastic modulus were, uterus (0.02–1.4 MPa), cervix (2.17–243 kPa), cardinal ligament (0.5–5.4 MPa), round ligament (9.1–14.0 MPa) and uterosacral ligament (0.75–29.8 MPa).

### **MSM Implementation Details**

We implemented deformable object with interconnected mass-spring-damper (MSD) system. The Young's Modulus is used as 5 for the uterus and Pois-

son's ratio is used as 0.49 for the uterus from [122]. Simulating deformable but cohesive bodies using a system of springs and dampers interconnected by point masses allows the real-world inter-atomic forces that give materials their elastic properties to be modelled on a macroscopic scale [119]. An MSD model provides a conceptually simple way to turn a 3D shape into an object which reacts in a visually believable manner to external forces. However, when the method is considered for an scientific application its use becomes more complex as the elastic properties of the MSD system now need to correlate with the materialistic properties input in to the simulation. A primary question therefore is how to translate a set of spring constants and damping coefficients such that they are scaled correctly to individual connections within an MSD so the entire system behaves according to the initial governing values [79].

The mathematics behind a typical MSD model is relatively simple, the spring mechanics are based in Newton's second law via Hooke's law ( $F = kx$ ) and damping is a function of the instantaneous velocity difference between two particles ( $b|v|$ ). This can be expressed in a form to connect two discrete points to each other as,

$$F_{ij} = (-k(|x| - d_{ij})\frac{\vec{x}}{|x|}) - (b|v|) \quad (7.2)$$

where  $F_{ij}$  is the force exerted on particle  $i$  due to the spring constant  $k$ ,  $d_{ij}$  the rest length of the spring between the particles,  $\vec{x}$  the vector representing the spring,  $b$  the damping coefficient and  $|v|$  the difference in current velocity between the particles.

### 7.1.2 Application of Generative Adversarial Networks

The Generative Adversarial Network (GAN) has shown tremendous potential in the machine learning world to create realistic-looking images and videos. GANs mainly generate amazing photorealistic images that mimic the content of datasets they were trained to replicate. One concern in medical imaging is whether GANs can also be effective in generating workable medical data as they are for generating realistic RGB images. In [124], the author proposed various GAN architectures from basic to more sophisticated style-based GANs on three medical imaging modalities and organs namely: cardiac cine-MRI, liver CT and RGB retina images. The top-performing GANs can generate realistic-looking medical images. However, segmentation results suggests that no GAN can reproduce the full richness of medical datasets [124].

A GAN is a system that consists of two models: a generator and a discriminator. The discriminator is simply a classifier that determines whether a given image is a real image from the dataset or an artificially generated image from the generator. This binary classifier will take the form of a convolutional neural network. The generator's task is to take in random input values (noise) and create an image from it using a deconvolutional neural network. The concept is like setting a seed for a random number generator — the same input noise will yield the same output. The generator uses the random noise as a seed of sorts to produce an image.

Image-to-image translation is an image synthesis task that requires the generation of a new image that is a controlled modification of a given image. Examples of image-to-image translation include:

- Translating summer landscapes to winter landscapes (or the reverse)
- Translating paintings to photographs (or the reverse)

- Translating horses to zebras (or the reverse)

Traditionally, training an image-to-image translation model requires a dataset comprised of paired examples. That is, a large dataset of many examples of input images  $X$  (e.g. summer landscapes) and the same image with the desired modification that can be used as an expected output image  $Y$  (e.g. winter landscapes). The requirement for a paired training dataset is a limitation. These datasets are challenging and expensive to prepare, e.g. photos of different scenes under different conditions. There is a desire for techniques for training an image-to-image translation system that does not require paired examples.

### **The uniqueness of our dataset**

The main challenging part of our dataset is the huge dissimilarity between the paired images (pre-BT and at-BT MRIs) which is an inherent property in brachytherapy. For this reason, it is extremely difficult for GANs as well to predict the anatomical deformation caused by an external force (i.e. the applicator) as there is a huge anatomical difference between pre-BT and at-BT MRIs. The CycleGAN approach is presented with many impressive applications:

- Style Transfer: It refers to the learning of artistic style from one domain, often paintings, and applying the artistic style to another domain, such as photographs.
- Object Transfiguration: It refers to the transformation of objects from one class, such as zebras into another class of objects, such as horses.
- Season Transfer: It refers to the translation of photographs taken in one season, such as summer, to another season, such as winter.

- MRI to CT: It refers to the transformation from MRI to CT image of the same patient where there is not much dissimilarity between the images.

There is inter-patient anatomical dissimilarity and extreme inpatient uterine deformation from pre-BT to at-BT in our dataset which makes the problem very unique and tough to address.

The main goal of GANs is to learn mapping functions between two domains  $X$  and  $Y$  given training samples  $\{x_i\}_{i=1}^N \in X$  and  $\{y_j\}_{j=1}^M \in Y$ . The model includes two mappings  $G : X \mapsto Y$  and  $F : Y \mapsto X$ . In addition, we introduce two adversarial discriminators  $D_X$  and  $D_Y$ , where  $D_X$  aims to distinguish between images  $\{x\}$  and translated images  $F(y)$ ; in the same way,  $D_Y$  aims to discriminate between  $\{y\}$  and  $\{G(x)\}$ . The objective contains kinds of two terms: adversarial losses for matching the distribution of generated images to the data distribution in the target domain; and a cycle consistency loss to prevent the learned mappings  $G$  and  $F$  from contradicting each other.

## Implementation details of CycleGAN

The Cycle Generative Adversarial Network, or CycleGAN, is an approach to training a deep convolutional neural network for image-to-image translation tasks. The Network learns mapping between input and output images using paired/unpaired dataset.

In our case we have two domains (pre-BT MRIs and at-BT MRIs). The model architecture is comprised of two generator models: one generator (Generator-A) for generating images for the first domain (Domain-A i.e. pre-BT) and the second generator (Generator-B) for generating images for the second domain (Domain-B i.e. at-BT).

$$\text{Domain} - B \rightarrow \text{Generator} - A \rightarrow \text{Domain} - A$$

$$\text{Domain} - A \rightarrow \text{Generator} - B \rightarrow \text{Domain} - B$$

Each generator has a corresponding discriminator model (Discriminator-A and Discriminator-B). The discriminator model takes real images from Domain and generated images from Generator to predict whether they are real or fake.

$Domain - A \rightarrow Discriminator - A \rightarrow [Real/Fake]$

$Domain - B \rightarrow Generator - A \rightarrow Discriminator - A \rightarrow [Real/Fake]$

$Domain - B \rightarrow Discriminator - B \rightarrow [Real/Fake]$

$Domain - A \rightarrow Generator - B \rightarrow Discriminator - B \rightarrow [Real/Fake]$

The loss used to train the Generators consists of several parts:

Adversarial Loss: Adversarial Loss to both the Generators, where the Generator tries to generate the images of its domain, while its corresponding discriminator distinguishes between the translated samples and real samples. Generator aims to minimize this loss against its corresponding Discriminator that tries to maximize it.

Cycle Consistency Loss: It captures the intuition that if we translate the image from one domain to the other and back again we should arrive at where we started. Hence, it calculates the L1 loss between the original image and the final generated image, which should look same as original image.

Identity Loss: It encourages the generator to preserve the color composition between input and output. This is done by providing the generator an image of its target domain as an input and calculating the L1 loss between input and the generated images.

We input a 128x128x1 grayscale MRI image (pre-BT) into the CycleGAN to generate a 128x128x1 MRI image (at-BT). The training was performed using 200 epochs and a batch size of 6. All the experiments with the neural network were implemented using Keras API [95] with TensorFlow and on two NVIDIA GeForce GTX 1080 GPU processors. It took around 6 hours to train the CycleGAN.

Figure 7.3, shows the result of CycleGAN on our dataset. From the visual results, it is clear that CycleGAN was not able to predict the at-BT MRI from pre-BT MRI due to the extreme organ deformations in brachytherapy.

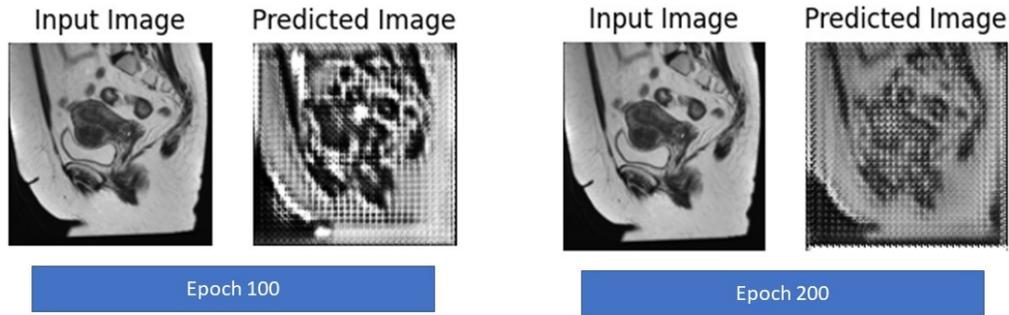


Figure 7.3: Results of CycleGAN in predicting at-BT MRI from pre-BT MRI

One shall keep in mind that training a GAN is often computationally intensive (typically because it involves two or more networks) and require a large amount of memory. Also, training GANs requires a lot of hyperparameter tuning which may or may not lead to better results [124]. This also affects more sophisticated GANs which, despite their good performances which improve the accuracy of the tasks, require large computing resources to train [124]. The results show that, most of the images generated by the tested GANs fail in reaching the baseline performance. A considerable amount of the medical data is acquired in a 3D fashion and voxel wise to achieve better performance. Typical GANs might not capture the full extent of the medical information when trained solely on 2D views. Indeed, this makes exploring GANs specially made for medical data an interesting research venue and could lead to an improvement in quality and ultimately clinical usability.

## 7.2 Conclusion

This study was performed on the dataset from 120 cervical cancer patients to predict the organ motion induced by the applicator insertion in brachytherapy. Large anatomical displacements and deformations were observed for most patients due to the insertion of the BT applicator between the anatomies before and at the time of BT. Standard DIR methods appeared inadequate to estimate such deformations. Our proposed CNN with autoencoder method achieved an average Dice Coefficient of 94.8% and a Hausdorff distance of 3.06 mm, in the automated uterus segmentation task. After that, a pre-trained modified U-net is proposed to predict the at-BT uterus position from only the pre-BT MRI. This method yielded an average Dice score of 89.5% and a Hausdorff distance of 3.6 mm in automatically predicting the uterine deformation due to applicator insertion. The applicator insertion induced the most significant mean displacement for the uterus of 62.5 mm. In addition, an average 27.8 mm displacement of the uterus has been measured between all the patients.

During this deformation procedure, we also measured the amount of deformation that happened to the canal shapes. As a result, the applicator insertion induced the average displacement of the uterine and vaginal canal of 43.2 mm and 22.1 mm, respectively. These deformation measurements of the canals give us an idea of how much the uterus could have been deformed based on these canal movements due to the applicator's presence. Based on these measurements, one can predict the at-BT uterus position, which will assist the clinicians in selecting the correct applicator size for each patient and help optimize personalized BT planning.

Predicting the deformed anatomy following applicator insertion from the natural anatomy before brachytherapy treatment is challenging as the uterus

and the other organs deform uniquely for each patient, depending on the patient’s anatomy. But the applicator model and its geometry is fixed for each patient. Our proposed method can deform the uterus shape from pre-BT to at-BT. As this is a shape-based deformable registration method, this method can be applied to other applications of non-rigid shape registration of various organs in MRI where the organs are delineated/appropriately segmented. In this work, we can deform the pre-BT uterine and vaginal canal to the applicator shape. This deformation field and the deformation parameters associated with this will be saved. After that, the pre-BT uterus shape will be deformed using the deformation of the uterine canal. In that way, we will be able to predict the deformed anatomy of the uterus employing just the pre-BT image.

The task of predicting the deformed anatomy following applicator insertion from the natural anatomy prior to brachytherapy treatment is very challenging as the uterus and the other organs deform uniquely for each patient, depending on the patient’s anatomy. The way the deformation takes place is also patient-specific. But the applicator model and its geometry is fixed for each patient. Our proposed method can deform the uterus shape from pre-BT to at-BT. As this is a shape-based deformable registration method, this method can be applied to other applications of non-rigid shape registration of various organs in MRI where the organs are delineated/appropriately segmented. Our novel DL-based algorithm, developed using 120-paired cervical cancer patient MR images, can quantify and predict the uterine deformation induced by applicator insertion. We are predicting this deformation from the natural anatomy before BT is challenging as the uterus and surrounding organs deform uniquely relative to each other and for every individual patient. This method demonstrates a first milestone towards achieving these technical BT parameters with improved accuracy and ease. No study has deployed DL in a

concerted effort to predict MR-based anatomical deformation in LACC BT to facilitate improved dosimetry. Analytics from this study can form a new kind of decision-aid, used before BT, for more efficient treatment customization, delivering personalized brachytherapy for women with LACC.

### **7.2.1 To Develop User Friendly Software/Library**

The digitization of modern imaging has led radiologists to become very familiar with computers and their user interfaces (UI). It is also known as the human-machine interface and defined as all the mechanisms (hardware or software) that supply information and commands to a user in order to accomplish a specific task within an interactive system. In practice, the UI is the link between the machine and the operator. In informatics, the UI includes inputs and outputs. Inputs communicate a user's needs to the machine, and the most common are the keyboard, mouse. The shift from analog to digital imaging should have led to an increase in efficiency among radiologists by reducing the time for interpretation and image manipulation. The proposed method for predicting at-BT uterine shape and location only from pre-BT MRI, will be implemented as an user-friendly software. This automated software will take the pre-BT MRI as the input and will produce the at-BT uterus. We will archive all the codes, pre-trained deep learning model weights for further training. Even if there is more data available, the model can be trained again with the new dataset to improve the performance accuracy. This software library can be used for automatic organ segmentation, image registration and organ deformation prediction as well.

User Interface (UI) Design focuses on anticipating what users might need to do and ensuring that the interface has elements that are easy to access, understand, and use to facilitate those actions. UI brings together concepts

from interaction design, visual design, and information architecture.

Steps to develop user-friendly library: Need to keep the interface simple.

- Input the pre-BT DICOM image for a new patient
- Pre-process the input image accordingly
- Load the pre-trained CNN models and the weights associated with it
- Segment the uterus automatically using CNN
- Predict the at-BT uterus shape and position using U-Net
- Measure the amount of uterine deformation using Free Form Deformation Model

Technology is changing every aspect of our lives including the ever-evolving world of medicine. It has enabled physicians to collect data in a more systematic way, explore different treatment methods, and find new tools to practice medicine. Doctor-patient communication has improved significantly – both through education (as doctors can use tools and 3D images to show patients what’s happening). Medical software is vital to the healthcare industry since it allows healthcare providers to monitor and manage organization and patient data efficiently. Our overall goal is to improve Locally Advanced Cervical Care Brachytherapy treatment and enhance clinical decision making and implant strategies prior to Brachytherapy for more efficient treatment customization and delivering personalized brachytherapy for women with LACC.

## 7.3 Future Work

GANs have achieved impressive results in image generation, image editing, and representation learning. Recent methods adopt the same idea for conditional image generation applications, such as text2image, image inpainting, and future prediction, as well as to other domains like videos and 3D models. The key to GANs' success is the idea of an adversarial loss that forces the generated images to be, in principle, indistinguishable from real images. This is particularly powerful for image generation tasks, as this is exactly the objective that much of computer graphics aims to optimize. We adopt an adversarial loss to learn the mapping such that the translated image cannot be distinguished from images in the target domain.

One of the essential components of a virtual reality surgical simulation is deformation. Deformations in computer graphics and surgical simulations are commonly modeled with three different approaches e.g. geometry based methods, Finite Element method (FEM), and Mass-spring Method (MSM). These methods take the physics of deformation into consideration. Even though the FEM results in more physically realistic deformations, a significant drawback of this method is its expensive computation cost and vulnerability to surgical procedures such as incision. However, MSM is relatively computationally inexpensive. Because of the real-time physics based behavior of MSM, it is widely accepted by surgical simulation community. For MSM, it is difficult to extract parameters from experiments for the thousands of individual springs and masses or dampers. To have both real-time performance and an acceptable deformation output, MSM is common choice for deformation simulations. To apply this scientifically, a 3D simulation model is necessary. Hence, in future, we will use 3D MRI datasets along with the organ segmentation to explore this model.

Our DL-based approach can also be used in other organ deformation tasks if we have enough training data. Also, this method could be generalized from 2D to 3D MRI datasets. However, we also need the corresponding uterus segmentation for the 3D dataset. If we have 3D pre-BT at-BT MRIs, along with the 3D uterus segmentation, mass-spring methods can be applied and the simulation model of the uterus can be created. In future, we will consider solving the problem of automated multi-organ segmentation from the 3D pelvic MRI dataset. After BT treatment, the multi-organ (tumour target, bladder, rectum, sigmoid, bowel) segmentation is performed manually to calculate the dose distribution. Currently the use of deep learning approaches in medical image analysis stay hindered by the limited access to huge annotated dataset. In contrast to EBRT and its technological advancements including contribution of AI, BT is still dependent more on the skills and technique of the physician than technological advances. There is a huge potential for incorporation of AI in BT technology.

# Bibliography

- [1] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, and Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*, 68(6):394–424, 2018.
- [2] Government of Canada Canadian Cancer Society. *Cervical cancer statistics*. <https://cancer.ca/en/cancer-information/cancer-types/cervical/statistics>.
- [3] Viswanathan AN and Beriwal S et al. American brachytherapy society consensus guidelines for locally advanced carcinoma of the cervix. part II: high-dose-rate brachytherapy. *Brachytherapy*, 11(1):47–52, 2012.
- [4] Banerjee R and Kamrava M. Brachytherapy in the treatment of cervical cancer: A review. *Int J Womens Health*, 6:555–564, 2014.
- [5] Green JA and Kirwan JM et al. Survival and recurrence after concomitant chemotherapy and radiotherapy for cancer of the uterine cervix: a systematic review and meta-analysis. *Lancet*, 358(9284):781–6, 2001.
- [6] Lanciano RM, Won M, Coia LR, and Hanks GE. Pretreatment and treatment factors associated with improved outcome in squamous cell carcinoma of the uterine cervix: A final report of the 1973 and 1978

- patterns of care studies. *Int J Radiat Oncol Biol Phys*, 20(4):667–676, 1991.
- [7] Potter R and Tanderup K et al. The EMBRACE II study: The outcome and prospect of two decades of evolution within the GEC-ESTRO GYN working group and the embrace studies. *Clinical and Translational Radiation Oncology*, 9:48–60, 2018.
- [8] Vale C, Tierney J, and Stewart L. Concomitant chemoradiotherapy for cervical cancer: a systematic review and meta-analysis of individual patient data. *Gynecol Oncol*, 100:442–443, 2006.
- [9] Han K, Milosevic M, and Fyles A et al. Trends in the utilization of brachytherapy in cervical cancer in the united states. *Int J Radiat Oncol Biol Phys*, 87:111–9, 2013.
- [10] Gill BS, Lin JF, and Krivak TC et al. National cancer data base analysis of radiation therapy consolidation modality for cervical cancer: the impact of new technological advancements. *Int J Radiat Oncol Biol Phys*, 90:1083–90, 2014.
- [11] Sonali S. Patankara and Ana I. Tergas et al. High versus low-dose rate brachytherapy for cervical cancer. *Gynecologic Oncology*, 136(3):534–541, 2015.
- [12] Prasert Lertsanguansinchai et al. Phase III randomized trial comparing ldr and hdr brachytherapy in treatment of cervical carcinoma. *International Journal of Radiation Oncology*, 59(5):1424–1431, 2004.
- [13] Bastien Rigaud and Guillaume Cazoulat et al. Modeling complex deformations of the sigmoid colon between external beam radiation therapy

- and brachytherapy images of cervical cancer. *Int J Radiation Oncol Biol Phys*, 106(5):1084–1094, 2020.
- [14] B Rigaud and A Klopp et al. Deformable image registration for dose mapping between external beam radiotherapy and brachytherapy images of cervical cancer. *Physics in Medicine and Biology*, 64(11), 2019.
- [15] Rigaud B, Simon A, and Castelli J et al. Deformable image registration for radiation therapy: Principle, methods, applications, and evaluation. *Acta Oncology*, 58:1225–1237, 2019.
- [16] Chetty IJ and Rosu-Bubulac M. Deformable registration for dose accumulation. *Semin Radiat Oncology*, 29:198–208, 2019.
- [17] G. E. Christensen and B. Carlson et al. Image-based dose planning of intracavitary brachytherapy: Registration of serial-imaging studies using deformable anatomic templates. *Int. J. Radiat. Oncol., Biol., Phys*, 51:227–243, 2001.
- [18] E. M. Vásquez Osorio and M. S. Hoogeman et al. A novel flexible framework with automatic feature correspondence optimization for nonrigid registration in radiotherapy. *Medical Physics*, 36:2848–2859, 2009.
- [19] L. Bondar and M. S. Hoogeman et al. A symmetric nonrigid registration method to handle large organ deformations in cervical cancer patients. *Medical Physics*, 37:3760–3772, 2010.
- [20] Eliana M. Vásquez Osorio and Inger-Karine K. Kolkman-Deurloo et al. Improving anatomical mapping of complexly deformed anatomy for external beam radiotherapy and brachytherapy dose accumulation in cervical cancer. *Medical Physics*, 42(1), 2015.

- [21] Kailyn Stenhouse et al. Development of a machine learning model for optimal applicator selection in high-dose-rate cervical brachytherapy. *Front. Oncol.*, 2021.
- [22] Xun Jia and Kevin Albuquerque. Artificial intelligence and deep learning for brachytherapy. *Seminars in Radiation Oncology*, 2022.
- [23] Susovan Banerjee et al. Artificial intelligence in brachytherapy: a summary of recent developments. *Br J Radiol*, 2021.
- [24] Tucker J Netherton et al. The emergence of artificial intelligence within radiation oncology treatment planning. *Oncology*, 2021.
- [25] Shrimanti Ghosh, Pierre Boulanger, Scott T Acton, Silvia S Blemker, and Nilanjan Ray. Automated 3D muscle segmentation from MRI data using convolutional neural network. *IEEE International Conference on Image Processing (ICIP)*, pages 4437–4441, 2017.
- [26] Shrimanti Ghosh, Nilanjan Ray, Pierre Boulanger, Kumaradevan Punithakumar, and Michelle Noga. Automated left atrial segmentation from magnetic resonance image sequences using deep convolutional neural network with autoencoder. *International Symposium on Biomedical Imaging*, 2020.
- [27] J. B. Maintz and M. A. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2:1–36, 1998.
- [28] W. R. Crum, T. Hartkens, and D. L. Hill. Non-rigid image registration: Theory and practice. *Br. J. Radiology*, 77:S140–S153, 2004.

- [29] E. M. Kerkhof and B. W. Raaymakers et al. Online mri guidance for healthy tissue sparing in patients with cervical cancer: An imrt planning study. *Radiotherapy and Oncology*, 88:241–249, 2008.
- [30] L. Xiong and A. Viswanathan. Deformable structure registration of bladder through surface mapping. *Medical Physics*, 33:1848–1856, 2006.
- [31] A. Godley and E. Ahunbay et al. Automated registration of large deformations for adaptive radiation therapy of prostate cancer. *Medical Physics*, 36:1433–1441, 2009.
- [32] Swamidas J and Kirisits C et al. Image registration, contour propagation and dose accumulation of external beam and brachytherapy in gynecological radiotherapy. *Radiother Oncology*, (<https://doi.org/10.1016/j.radonc.2019.08.023>), 2020.
- [33] Balsiger F, Steindel C, Arn M, Wagner B, Grunder L, and El-Koussy M et al. Segmentation of peripheral nerves from magnetic resonance neurography: a fully-automatic, deep learning-based approach. *Front Neurol*, 2018.
- [34] Nicolae A, Morton G, Chung H, Loblaw A, Jain S, and Mitchell D et al. Evaluation of a machine-learning algorithm for treatment planning in prostate low-dose-rate brachytherapy. *Int J Radiat Oncol Biol Phys*, 2017.
- [35] Stenhouse K, Ciunkiewicz P, Roumeliotis M, and McGeachy P. Use of machine learning algorithms to identify predictive geometric features for optimal applicator selection in high dose-rate (hdr) cervical brachytherapy: scientific session 5b: Brachytherapy. *Medical Physics*, 2019.

- [36] Meyer P et al. Survey on deep learning for radiotherapy. *Computers in Biology and Medicine*, 98:126–146, 2018.
- [37] William W, Ware A, Basaza-Ejiri A.H., and Obungoloch J. A review of image analysis and machine learning techniques for automated cervical cancer screening from pap-smear images. *Computer Methods and Programs in Biomedicine*, 164:15–22, 2018.
- [38] Cunha J.A.M, Flynn R, and Bélanger C et al. Brachytherapy future directions. *Seminars in Radiation Oncology*, 30:94–106, 2020.
- [39] Banerjee S, Kataria T, Gupta D, Goyal S, Bisht SS, and Basu T et al. Use of ultrasound in image- guided high- dose- rate brachytherapy: enumerations and arguments. *J Contemp Brachytherapy*, 2017.
- [40] Whitten P. Artificial intelligence driven diagnosis of lung cancer in patients with multiple pulmonary nodules. *Chest*, 2019.
- [41] Daguang Zhang et al. Automatic segmentation and applicator reconstruction for ct-based brachytherapy of cervical cancer using 3d convolutional neural networks. *J. Appl. Clin. Med. Phys*, 2020.
- [42] Lei Y, Tian S, and He X et al. Ultrasound prostate segmentation based on multidirectional deeply supervised v-net. *Medical Physics*, 46:3194–3206, 2019.
- [43] Gessert N, Priegnitz T, and Saathoff T et al. Spatio-temporal deep learning models for tip force estimation during needle insertion. *International Journal of Computer Assisted Radiology and Surgery*, 14:1485–1493, 2019.

- [44] Wang F, Xing L, Bagshaw H, Buyyounouski M, and Han B. Deep learning applications in automatic needle segmentation in ultrasound-guided prostate brachytherapy. *Medical Physics*, 2020.
- [45] De Leeuw AA et al. Applicator reconstruction and applicator shifts in 3d mr-based pdr brachytherapy of cervical cancer. *Radiother Oncol*, 2009.
- [46] T. Nagashima, T. Shirakuni, and S. I. Rapoport. A two-dimensional, finite element analysis of vasogenic brain edema. *Neurologia Medico-Chirurgica*, 30(1):1–9, 1990.
- [47] K. K. Mendis, R. Stalnaker, and S. H. Advani. A constitutive relationship for large-deformation finite-element modeling of brain-tissue. *J. Biomech. Eng. Trans. ASME*, 117:279–285, 1995.
- [48] Sharp G, Fritscher KD, and Pekar V et al. Vision 20/20: Perspectives on automated image segmentation for radiotherapy. *Med Phys*, 41, 2014.
- [49] Han X, Hoogeman MS, and Levendag PC et al. Atlas-based auto segmentation of head and neck CT images. *Med Image Comput Comput Assist Interv*, 11:434–441, 2008.
- [50] Cardenas C, Wong A, and Mohamed A et al. Delineating high-dose clinical target volumes for head and neck tumors using machine learning algorithms. *Med Phys*, 2016.
- [51] Cardenas CE, McCarroll R, and Court LE et al. Deep learning algorithm for auto-delineation of high-risk oropharyngeal clinical target volumes with built-in dice similarity coefficient parameter optimization function. *Med Phys*, 44:3160–3161, 2017.

- [52] Carlos E Cardenas and Rachel E McCarroll et al. Deep learning algorithm for auto-delineation of high-risk oropharyngeal clinical target volumes with built-in dice similarity coefficient parameter optimization function. *Int J Radiat Oncol Biol Phys*, 101(2):468–478, 2018.
- [53] Elhawary H et al. Multimodality non-rigid image registration for planning, targeting and monitoring during ct-guided percutaneous liver tumor cryoablation. *Acad Radiol*, 2010.
- [54] Ganser KA et al. A deformable digital brain atlas system according to talairach and tournoux. *Med Image Anal*, 2004.
- [55] Foskey M et al. Large deformation threedimensional image registration in image-guided radiation therapy. *Phys Med Biol*, 2005.
- [56] Dornheim L et al. Automatic segmentation of the left ventricle in 3d spect data by registration with a dynamic anatomic model. *MICCAI*, 2005.
- [57] Grosland NM, Bafna R, and Magnotta VA. Automated hexahedral meshing of anatomic structures using deformable registration. *Comput Methods Biomech Biomed Eng*, 2009.
- [58] G. E. Christensen and H. J. Johnson. Consistent image registration. *IEEE Transactions on Medical Imaging*, 2001.
- [59] R. Stefanescu, X. Pennec, and N. Ayache. Grid powered nonlinear image registration with locally adaptive regularization. *Medical Image Analysis*, 2004.
- [60] Kun Han et al. Diffeomorphic image registration with neural velocity field. *CVPR*, 2022.

- [61] Huilin Yang et al. Lddmm-face: Large deformation diffeomorphic metric learning for flexible and consistent face alignment. *CVPR*, 2021.
- [62] Ali Khan, Elizabeth Aylward, Patrick Barta, Michael Miller, and M. Faisal Beg. Semi-automated basal ganglia segmentation using large deformation diffeomorphic metric mapping. *MICCAI*, 2005.
- [63] Ali Khan, Lei Wang, and M. Faisal Beg. Multistucture large deformation diffeomorphic brain registration. *IEEE Transactions on Biomedical Engineering*, 2013.
- [64] E Basafa and F Farahmand. Real-time simulation of the nonlinear viscoelastic deformations of soft tissues. *International Journal of Computer Assisted Radiology and Surgery*, 6:297–307, 2011.
- [65] T Halic et al. Soft tissue deformation and optimized data structures for mass spring methods. *Proceedings of the 2009 Ninth IEEE International Conference on Bioinformatics and Bioengineering, BIBE*, pages 45–52, 2009.
- [66] C Paloc, F Bello, R Kitney, and A Darzi. Online multiresolution volumetric mass spring model for real time soft tissue deformation. *Proceedings of the 5th International Conference on Medical Image Computing and Computer-Assisted Intervention-Part II, MICCAI*, page 219–226, 2002.
- [67] K Miller, A Wittek, and G Joldes. Biomechanics of the brain for computer-integrated surgery. *Acta of Bioengineering and Biomechanics*, 12:25–37, 2010.
- [68] G. R. Joldes, A. Wittek, and K. Miller. Real-time nonlinear finite element computations on gpu - application to neurosurgical simulation. *Com-*

- puter Methods in Applied Mechanics and Engineering*, 199:3305–3314, 2010.
- [69] Plantefeve R, Peterlik I, Haouchine N, and Cotin S. Patient specific biomechanical modeling for guidance during minimally invasive hepatic surgery. *Ann Biomed Eng*, 44:139–53, 2016.
- [70] Miller K, Joldes G, Lance D, and Wittek A. Total lagrangian explicit dynamics finite element algorithm for computing soft tissue deformation. *Communications in Numerical Methods in Engineering*, 23:121–34, 2007.
- [71] Jinao Zhang, Yongmin Zhong, and Chengfan Gu. Neural network modelling of soft tissue deformation for surgical simulation. *Artificial Intelligence in Medicine*, 97:61–70, 2019.
- [72] Yuping Duan and Weimin Huang et al. Modeling and simulation of mass-spring model for soft tissue deformation. *MICCAI*, 11, 2012.
- [73] A Nealen et al. Physically based deformable models in computer graphics. *Comput. Graph. Forum*, 25:809–836, 2006.
- [74] M Muller et al. Point based animation of elastic, plastic and melting objects. *ACM SIGGRAPH Symposium on Computer Animation*, 2004.
- [75] O Zienkiewicz et al. The finite element method: Its basis and fundamentals. *The Finite Element Method, Elsevier Butterworth-Heinemann*, 2005.
- [76] F Chen et al. Soft tissue modeling using non-linear mass spring and simplified medial representation. *Annu Int Conf IEEE Eng Med Biol Soc*, 2007.

- [77] D Zerbato et al. Calibration of mass spring models for organ simulations. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, page 370–375, 2007.
- [78] E Basafa, F Farahmand, and G Vossoughi. A non-linear mass-spring model for more realistic and efficient simulation of soft tissues surgery. *Stud. Health Technol. Inform.*, 132:23–25, 2008.
- [79] B A Lloyd et al. Identification of spring parameters for deformable object simulation. *IEEE Trans. Vis. Comput. Graph.*, 13:1081–1094, 2007.
- [80] G San-Vicente et al. Cubical mass-spring model design based on a tensile deformation test and nonlinear material model. *IEEE Trans. Vis. Comput. Graph.*, 18:228–241, 2012.
- [81] Chen J et al. Investigating rectal toxicity associated dosimetric features with deformable accumulated rectal surface dose maps for cervical cancer radiotherapy. *Radiation Oncology*, 13(125), 2018.
- [82] Brock K et al. Accuracy of finite element model-based multi-organ deformable image registration. *Med Phys*, 32:1647–1659, 2005.
- [83] Thirion J P. Image matching as a diffusion process: an analogy with maxwell’s demons. *Med. Image Anal.*, 2:243–260, 1998.
- [84] Noriyuki Kadoya et.al. Evaluation of deformable image registration between external beam radiotherapy and HDR brachytherapy for cervical cancer with a 3d-printed deformable pelvis phantom. *Med Phys*, 44(4), 2017.
- [85] Mattes et al. Nonrigid multimodality image registration. *SPIE Medical Imaging*, 4322:1609–1620, 2001.

- [86] Qiuhong Ke, Jun Liu, and Mohammed Bennamoun et al. Computer vision for human-machine interaction. *Computer Vision for Assistive Healthcare*, 2018.
- [87] McCormick M, Liu X, Jomier J, Marion C, and Ibanez L. Itk: enabling reproducible research and open science. *Front Neuroinform*, 2014.
- [88] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. *CVPR*, 2015.
- [89] Ronneberger O, Fischer P, and Brox T. U-net: Convolutional networks for biomedical image segmentation. *Med Image Comput Comput Assist Interv*, pages 234–241, 2015.
- [90] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.
- [91] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv:1412.6980*, 2014.
- [92] Wei Wang, Yan Huang, Yizhou Wang, and Liang Wang. Generalized autoencoder: A neural network framework for dimensionality reduction. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2014.
- [93] NiBabel. *Defining the DICOM orientation*  
. [https://nipy.org/nibabel/dicom/dicom\\_orientation.html](https://nipy.org/nibabel/dicom/dicom_orientation.html).
- [94] Xiaoran Zhang, David Glynn Martin, Michelle Noga, and Kumaradevan Punithakumar. Fully automated left atrium segmentation from anatom-

- ical cine long-axis mri sequences using deep convolutional neural network with unscented kalman filter. *Medical Image Analysis*, submitted, 2020.
- [95] Chollet Francois et al. Keras. *published by GitHub*, 2015.
- [96] Russakovsky O, Deng J, Su H, Krause J, and Satheesh S et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 2015.
- [97] Cheng JZ, Ni D, Chou YH, Qin J, and Tiu CM et al. Computer-aided diagnosis with deep learning architecture: Applications to breast lesions in us images and pulmonary nodules in ct scans. *Scientific Reports*, 2016.
- [98] Roth HR, Lu L, Liu J, Yao J, and Seff A et al. Improving computer-aided detection using convolutional neural networks and random view aggregation. *IEEE Transactions on Medical Imaging*, 2016.
- [99] Li R, Zhang W, Suk HI, Wang L, and Li J et al. Deep learning based imaging data completion for improved brain disease diagnosis. *Proceedings of Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2014.
- [100] Shin HC, Roth HR, Gao M, Lu L, and Xu Z et al. Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. *IEEE Transactions on Medical Imaging*, 2016.
- [101] Gupta A, Ayhan M, and Maida A. Natural image bases to represent neuroimaging data. *Proceedings of International Conference on Machine Learning (ICML)*, 2013.

- [102] Nie D, Wang L, Gao Y, and Sken D. Natural image bases to represent neuroimaging datafully convolutional networks for multi-modality isointense infant brain image segmentation. *Proceedings of IEEE International Symposium on Biomedical Imaging (ISBI)*, 2016.
- [103] Su H, Xing F, Kong X, Xie Y, Zhang S, and Yang L. Robust cell detection and segmentation in histopathological images using sparse reconstruction and stacked denoising autoencoders. *Proceedings of Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015.
- [104] Xu J, Xiang L, Liu Q, Gilmore H, and Wu J et al. Robust cell detection and segmentation in histopathological images using sparse reconstruction and stacked denoising autoencoders. *Proceedings of Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015.
- [105] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial net. *Advances in neural information processing systems*, 2014.
- [106] S Ioffe and C Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint*, 2015.
- [107] Nikhil Varma Keetha, Samson Anosh Babu P, and Chandra Sekhara Rao Annavarapu. U-det: A modified u-net architecture with bidirectional feature network for lung nodule segmentation. *Computer Vision and Pattern Recognition*, 2020.
- [108] Tan M and Pang R et al. Efficientdet: Scalable and efficient object detection. *arXiv preprint*, 2019.

- [109] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. *CVPR*, 2018.
- [110] Bengio Y and Grandvalet Y. No unbiased estimator of the variance of k-fold cross-validation. *Journal of machine learning research*, 2004.
- [111] He J and Christensen GE. Large deformation inverse consistent elastic image registration. *Inf Process Med Imaging*, 2003.
- [112] Le Guyader C and Vese LA. A combined segmentation and registration framework with a nonlinear elasticity smoother. *Computer Vision and Image Understanding*, 2011.
- [113] Lu X, Zhao Y, Zhang B, Wu JS, Li N, and Jia WT. A non-rigid cardiac image registration method based on an optical flow model. *Int J Light Electron Optics*, 2013.
- [114] Rueckert D, Sonoda LI, Hayes C, Hill D, Leach M, and Hawkes D. Non-rigid registration using free-form deformations: Application to breast mr images. *IEEE Trans Med Imaging*, 1999.
- [115] Rohde GK, Aldroubi A, and Dawant BM. The adaptive bases algorithm for intensity-based nonrigid image registration. *IEEE Trans Med Imaging*, 2003.
- [116] Mohammad Rouhani and Angel D. Sappa. Non-rigid shape registration: A single linear least squares framework. *ECCV*, pages 264–277, 2012.
- [117] Fujiwara K., Nishino K., Takamatsu J., Zheng B., and Ikeuchi K. Locally rigid globally non-rigid surface registration. *In: Proc. IEEE International Conference on Computer Vision*, page 1527–1534, 2011.

- [118] Rusinkiewicz S. and Levoy M. Efficient variants of the icp algorithm. *In: Proc. IEEE International Conference on on 3-D Digital Imaging and Modeling*, 2001.
- [119] S. M. Longshaw et al. Integration of spring physics with the sph method for quasi-solid to fluid interaction using gpgpu programming. *SPHERIC Workshop*, 2013.
- [120] Ostoja-Starzewski et al. Lattice models in micromechanics. *Appl. Mech*, 2002.
- [121] Gurpreet Singh and Arnab Chanda. Mechanical properties of whole-body soft human tissues: a review. *Biomedical Materials*, 2021.
- [122] Dias N et al. Pelvic floor dynamics during high-impact athletic activities: a computational modeling study. *Clin. Biomech*, 2017.
- [123] Tan T Baah-Dwomoh A, McGuire J and de Vita R. Mechanical properties of female reproductive organs and simulants for sports impact surrogates. *J. Mech. Behav. Biomed. Mater*, 2016.
- [124] Youssef Skandarani et al. Gans for medical image synthesis: An empirical study. *arXiv:2105.05318*, 2021.