Towards Personalized PTV Margins for External Beam Radiation Therapy of the Prostate

by

Andrew Coathup
BSc, University of Ottawa, 2014

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Physics and Astronomy

Towards Personalized PTV Margins for External Beam Radiation Therapy of the Prostate

by

Andrew Coathup
BSc, University of Ottawa, 2014

Supervisory Committee

Dr. P. Basran, Co-Supervisor
(Department of Physics and Astronomy)

Dr. M. Bazalova-Carter, Co-Supervisor
(Department of Physics and Astronomy)

Dr. W. Ansbacher, Departmental Member
(Department of Physics and Astronomy)

Supervisory Committee

Dr. P. Basran, Co-Supervisor
(Department of Physics and Astronomy)

Dr. M. Bazalova-Carter, Co-Supervisor
(Department of Physics and Astronomy)

Dr. W. Ansbacher, Departmental Member
(Department of Physics and Astronomy)

Abstract

External Beam Radiation Therapy (EBRT) is a common treatment option for patients with prostate cancer. When treating the prostate with EBRT, a geometric volume (PTV margin) is added around the prostate to account for uncertainties in treatment planning and delivery. Current methods for estimating PTV margins rely on the analysis of population-based inter- and intra-fraction motion data. These methods do not consider the patient-to-patient differences in demographic or clinical presentation of patient specific factors (PSFs), such as age, weight, body-mass index, health and performance status, prostate-specific antigen levels, Gleason scores, presence of bowel problems, or other health conditions. The purpose of this thesis is to investigate the feasibility using regression-based predictive algorithms to predict the extent of prostate motion for the purpose of

personalizing the PTV margin using PSFs as inputs. Benchmarking simulations of Linear, Ridge, LASSO, SVR, kNN, and MLP algorithms were performed by simulating prostate intra-fraction motion and realistic variations in PSFs. Sample sizes ranged from n=20 to 800, with varying levels of noise into the motion data (0-10mm). Leave-one-out cross validation was used to train and validate algorithm performance. The results suggest that algorithm performance improves significantly within the first 50 – 100 patients, and this rate of improvement is independent of noise in prostate motion. The Ridge regression algorithm predicted intra-fraction motion to the lowest mean absolute error in simulated motion, performing especially well in small datasets. To evaluate the clinical utility of this approach, pre- and post-treatment prostate motion data, treatment time data, and rectal distension data was recorded in 21 patients, along with a variety of PSFs. In the analysis of patient data, the LASSO algorithm out-performed the Ridge algorithm, predicting the mean and standard deviation of an individual prostate cancer patient's intra-fraction motion to within 0.8mm and 0.4 mm mean absolute error, respectively. However, prostate motion predictions did not correlate with PSFs, possibly due to the small sample size. This work demonstrates the feasibility of using regression-based algorithms for predicting prostate motion, and hence the opportunity to personalize PTV margins in prostate cancer patients.

Contents

Supervisory Committee ii
Abstract iii
Contents v
List of Tables vii
List of Figures ix
Acknowledgments x
1 Introduction 1
1.1 Radiation Therapy 1
1.1.1 Volumes in Radiation Therapy 3
1.2 PTV Margin 4
1.2.1 Motivation for Personalized PTV Margin 4
1.3 Thesis Scope 5
2 Radiation Therapy Background 7
2.1 External Beam Radiation Therapy 7
2.1.1 Linac Photon Generation and Modulation 7
2.1.2 Linac Treatment Couch and Coordinate System 9
2.1.3 Linac on-board imaging 10
2.2 External Beam Radiation Therapy Process 11
2.2.1 External Beam Radiation Therapy of the Prostate 12
2.3 Volumes and Margins in Radiation Therapy 14
2.4 PTV Margin Estimation 15
2.4.1 Standard Deviation Approach 15
2.4.2 van Herk Formula 16
3 Statistical Analysis Background 19
3.1 Overview 19
3.2 Data Preprocessing 19
3.3 Basic Statistics Definitions 21
3.4 Predictive Algorithms 22
3.4.1 General Overview 22
3.4.2 The Predictive Algorithm Process 23
4 Methodology 29
4.1 Data Collection 29
4.1.1 Time Independent Data 30
4.1.2 Time Dependent Data 36
4.2 Data Analysis 38
4.2.1 Overview of the software environment 38
4.2.2 Data Preprocessing 38
4.2.3 Searching for Knowledge 40
4.3 Predictive Algorithms for Intra-Fraction Motion 43
4.3.1 Overview of Software Environment 43
4.3.2 Predictive Algorithm Selection 43
4.3.3 Training and Validation of Algorithm Performance 49
4.3.4 Benchmarking of Predictive Algorithms 50

4.3.5 Applying Predictive Algorithms to Real Patient Data 55
4.4 Generating a Patient Specific Margin 56
4.4.1 Population Based Margin 56
4.4.2 Personalized Margin 57
5 Results 58
5.1 Data Analysis 58
5.1.1 Patient Specific Factor Statistics 58
5.1.2 Identification of Patient Specific Factors 59
5.1.3 Time Dependent Data 66
5.1.4 Rectal Distention 66
5.1.5 Population-Based Statistics 70
5.2 Predictive Algorithm Benchmarking 72
5.2.1 Number of Patients Required 72
5.2.2 Evaluation of Algorithms 74
5.2.3 Influence of Noise 77
5.3 Predictive Algorithms Applied to Real Patient Data 80
5.3.1 Intra-fraction Motion 80
5.3.2 Algorithm Coefficient Weighting 81
5.4 Generating Margins 86
5.4.1 Population-Based Margin 86
5.4.2 Personalized Margin 87
6 Discussion 92
6.1 Analysis of Patient Specific Factors for Real Patients 92
6.2 Time Dependent Data 94
6.3 Rectal Distension 97
6.4 Population-based Statistics 100
6.5 Predictive Algorithm Benchmarking 101
6.6 Predictive Algorithms for Real Patient Data 102
6.7 PTV Margin Evaluation 104
7. Conclusion 106
7.1 Future Work 109
Bibliography 112
Appendix 116

List of Tables

Table 1: Gas permutation acronyms. 42
Table 2: Sampling range for each PSF 51
Table 3: Example of how incomplete data was filled for a PSF 55
Table 4: Estimation of errors 56
Table 5: PSF statistics 58
Table 6: Number of implanted fiducial markers (seeds) patient distribution 58
Table 7: ECOG status patient distribution 58
Table 8: Boolean PSF statistics 59
Table 9: Pearson and Spearman correlations (>0.3) between mean 3D intra-fraction motion and PSFs 60
Table 10: Pearson and Spearman correlations (>0.3) between the standard deviation of 3D intra-fraction motion and PSFs 61
Table 11: Pearson and Spearman correlation between mean 6D intra-fraction rotation and PSFs 62
Table 12: Pearson and Spearman correlation between the standard deviation of 6D intra-fraction rotation and PSFs 63
Table 13: Pearson and Spearman correlation between mean 6D inter-fraction rotation and PSFs 64
Table 14: Pearson and Spearman correlation between the standard deviation of 6D inter-fraction rotation and PSFs 65
Table 15: Correlation between vertical and longitudinal intra-fraction motion 66
Table 16: Effect of gas on mean intra-fraction 3D motion 67
Table 17: Effect of gas on the standard deviation of intra-fraction 3D motion 67
Table 18: Effect of pre-treatment gas on mean inter-fraction rotation 68
Table 19: Effect of pre-treatment gas on the standard deviation of inter-fraction rotation 68
Table 20: Effect of a change in gas during treatment on mean intra-fraction rotation 69
Table 21: Effect of a change in gas during treatment on the standard deviation of intra-fraction rotation 69
Table 22: Mean and standard deviation of intra-fraction motion for all 678 fractions 70
Table 23: Directionality of intra-fraction motion. Mean and standard deviation of intra-fraction motion along the positive and negative directions for each axis 70
Table 24: Mean and standard deviation of inter-fraction rotation for all 678 fractions ... 71
Table 25: Mean and standard deviation of intra-fraction rotation for all 678 fractions ... 71
Table 26: Directionality of patient rotation. Mean and standard deviation of inter-fraction rotation along the positive and negative directions for each axis 72
Table 27: Directionality of patient motion. Mean and standard deviation of intra-fraction rotation along the positive and negative directions for each axis 72
Table 28: PSF coefficients used in mean intra-fraction motion predictions for LASSO and Ridge models. 83
Table 29: PSF coefficients used in standard deviation of intra-fraction motion predictions for LASSO and Ridge models. 84

Table 30: Prediction error (MAE) comparison between LASSO and Ridge models for mean intra-fraction motion 84
Table 31: Prediction error (MAE) comparison between LASSO and Ridge models for standard deviation of intra-fraction motion 85
Table 32: Population-based PTV margin and errors..... 86
Table 33: Personalized PTV margin for one patient..... 88
Table 34: Relative volume receiving at least 70, 50, 30, and 20 Gy, respectively. 91
Table 35: Mean intra-fraction longitudinal motion between ECOG status 93
Table 36: Range of sampled timestamps and correlations with intra-fraction motion 95
Table 37: Occurrence of a change in gas broken down by individual patient. Patients with greater than 1 fractions with a gas change are highlighted in red. 99
Table 38: Reproducibility of 3D intra-fraction motion..... 105

List of Figures

Figure 1: Linac with important components (from Podgorsak, 2005, pg. 140)..... 8
Figure 2: Linac with accompanying coordinate system axes. 9
Figure 3: Linac OBI with kV x-ray source and detector labelled. 10
Figure 4: Example of five-field prostate EBRT plan. Five intensity-modulated radiation fields are delivered at gantry angles of 100°, 50°, 310°, 260°, and 0°. The red circle is the PTV contour. 13
Figure 5: 3D representation of prostate and neighbouring organs, along with motion and rotation axes. 13
Figure 6: Margins defined in radiation therapy in accordance with ICRU 62 (courtesy I Spadinger). 14
Figure 7: Projected performance of algorithms based on simulated data 73
Figure 8: Performance of the Linear model using simulated data from 20, 50, and 200 patients. 75
Figure 9: Performance of the Ridge model using simulated data from 20, 50, and 200 patients. 76
Figure 10: Performance of the MLP model using simulated data from 20, 50, and 200 patients. 77
Figure 11: Effect of noise on linear model 78
Figure 12: Effect of noise on ridge model 79
Figure 13: Effect of noise on the MLP model 79
Figure 14: LASSO algorithm prediction of a patient's mean intra-fraction motion..... 80
Figure 15: LASSO algorithm prediction of the standard deviation of a patient's intra-fraction motion 81
Figure 16: LASSO MAE and magnitude of PSF coefficients plotted against choice of alpha tuneable parameter 85
Figure 17: Comparison of original PTV and new PTV for one patient (red contours). The new PTV is on the left. The original PTV is on the right. 89
Figure 18: Dose distribution near important organs-at-risk. The bladder (green contour) and rectum (brown contour) are both shown here. The new PTV plan is on the left. The original PTV plan is on the right. The PTV (red contour) is visible in this slice on the original plan only. The PTV used for the new plan does not extend into this slice. 90
Figure 19: Dose-volume histogram (DVH) of PTV and OAR volumes under both the new and original plans. The new and original PTVs are irradiated to the same coverage. The OAR all receive lower dose when using the new PTV compared to the original PTV. ... 91
Figure 20: Scatterplot between BMI and standard deviation of intra-fraction longitudinal motion 94
Figure 21: Example of time dependent csv file input for one patient. 116
Figure 22: Example of time independent csv file input for one patient 116
Figure 23: Projected performance of algorithms based on simulated data (with error bars) 117
Figure 24: Effect of noise on linear model (with error bars) 117
Figure 25: Effect of noise on ridge model (with error bars) 118
Figure 26: Effect of noise on MLP model (with error bars) 118

Acknowledgments

Thank you to all the people at the BCCA and UVIC who made the past two years enjoyable! Thanks to Parminder, especially for the encouragement and patience! Thanks to Magdalena, Will, and Deidre for the helpful thesis comments! Thanks to Chelsea, Tom, and Kossivi for making courses more pleasant! Thanks to Susan, and then later, Sarah for being office mates at BCCA! Thanks to Pramodh, Paul, Nafisa, Lichen, Dylan, Kristy, and Tyler for being UVIC office mates – thankfully not all at the same time! Thanks to the Victoria swing dancing community for providing a place to go dancing and a non-academic social environment! Similarly, thanks to the beach and floor volleyball crews! Thanks to my dad for providing national-geographic-level emails about his various travels! Thanks to Brojar and Trevstar for being super cool brothers my whole life! Thanks to my mum for always being there when I need to talk! Biggest thanks to Ildara for being kind and supportive all of the time, and laughing at my jokes most of the time!

1 Introduction

1.1 Radiation Therapy

Radiation therapy is a cancer treatment option that uses ionizing radiation to damage and/or kill cancer cells in order to cease their growth. Radiation therapy is used with either curative or palliative intent. Curative radiation therapy is treatment with intent to permanently kill the cancer. Palliative radiation therapy is treatment with intent to limit symptoms of the cancer. Approximately 50% of patients diagnosed with cancer receive radiation therapy as a part of their treatment [1], possibly in conjunction with other treatments such as surgery and chemotherapy, making it a very common option for cancer care.

At the molecular level, radiation therapy is fundamentally a stochastic process. Ionizing radiation, defined as radiation of sufficient energy to ionize the medium, enters the tissue and deposits none, some, or all its energy [2]. The energy that is deposited typically comes in the form of electrons which are set into motion by the incident radiation, resulting in ionizations of molecules within the individual cells that make up the tissue [3]. The total energy deposited into the irradiated tissue can be described quantitatively by the absorbed dose, measured in Gy [J/kg], or the mean energy deposited per kg of tissue.

Deposited energy can cause radiation induced DNA damage to occur through two avenues. Radiation can either directly ionize the bonds within the DNA (direct damage) or it can ionize other components of the cell (typically water) producing free radicals, which

subsequently damage the cell's DNA [4]. DNA is a critically important component of the cell and damage to the DNA is lethal to the cell if that damage is not repairable. In order to improve the likelihood of killing the cancer cells and/or controlling the growth of the cancer body, it is important to irradiate the cancerous region to a sufficiently high and uniform dose. A larger dose increases the number of ionizations that occur within the cells, which leads to more DNA damage, which increases the probability of cellular death. A larger dose also increases the risk of normal tissue complications, however [4].

Normal tissue complications occur because normal (non-cancer) tissue cells are affected by radiation. The complications that arise due to this irradiation can be either short term (days to weeks) or long term (months to years) [4]. Great care is taken to deliver radiation in a way that achieves the goal of the treatment, while also minimizing normal tissue complications. This includes setting various maximum dose limits to radiation sensitive organs, often including volumetric dose limits as well. These limits are specific to each organ as there is variation in organ susceptibility to radiation. Given the necessity of irradiating the cancerous region to a sufficiently high and uniform dose, while also being mindful of the surrounding radiation sensitive normal tissues (organs), the differences between cancer cells and normal cells are important to exploit.

There are two main considerations when selectively targeting cancer cells, and not normal cells: biological considerations and geometric considerations. Biological considerations refer to the differences in biological response to radiation between normal cells and cancer cells. In general, cancer cells are not able to recover from sub-lethal radiation damage as

well as normal tissue cells. This property is exploited through fractionation. By irradiating the cells in small doses each day instead of a single large dose, the cumulative cell death is ultimately greater for the cancer cells relative to the normal tissue cells. Geometric considerations refer to the position of cancer in the body. Cancer destined for treatment with radiation is often quite localized. The cancer cells are typically clustered together, forming a tumour body. By selectively irradiating the region consisting of an overwhelming proportion of cancer cells, a larger proportion of the tumour cells are likely to die compared to the normal cells.

1.1.1 Volumes in Radiation Therapy

ICRU 62 defines several geometric volumes involved in the radiation therapy process [5]. These volumes are defined to guide treatment planning by specifying the geometric position of the cancerous region as well as the surrounding radiation sensitive normal tissues. While all ICRU treatment volumes are discussed in the radiation therapy background section, two important volumes relevant to this thesis are the CTV and PTV. The CTV is defined as a purely clinical volume independent of treatment modality, whereas the PTV is defined dependent on the chosen treatment modality and is the volume to which dose is prescribed and reported. Their ICRU definitions are included below, but also discussed in more detail in chapter two:

The clinical target volume (CTV) is defined in ICRU 62 as "*a tissue volume that contains a demonstrable GTV and/or subclinical malignant disease that must be eliminated. This volume must be treated adequately in order to achieve the aim of radical therapy."*

The planning target volume (PTV) is defined in ICRU 62 as *"a geometrical concept used for treatment planning, and it is defined to select appropriate beam sizes and beam arrangements, to ensure that the prescribed dose is actually delivered to the CTV."*

1.2 PTV Margin

The PTV margin is an additional geometric volume that surrounds the identified cancerous region (CTV). The size of the PTV margin is set to account for the spatial uncertainties introduced in the radiation therapy process and ensures the CTV actually receives the prescription radiation dose. Uncertainties in treatment planning, set-up, and delivery are all accounted for in the PTV margin [6].

Ideally the PTV margin is large enough to account for all the uncertainties previously mentioned, but small enough to limit significant dose to important normal tissue structures, and consequentially, the side effects of the treatment. This is especially important considering that the PTV margin, at least in principle, consists of only normal tissue cells which do not require radiation. In prostate cancer treatment particularly, the PTV margin may extend isotropically by several millimeters to include portions of the rectum and bladder. These two important organs are likely to suffer radiation toxicity.

1.2.1 Motivation for Personalized PTV Margin

PTV margins are currently estimated using population-based approaches. This may be performed by measuring the mean and standard deviation of tumour motion across a large number of patients and using these statistics to estimate the behaviour of future patients

[7]. Another, more specific approach, is to break down the individual contributions of the uncertainties in the planning, set-up, and delivery of radiation therapy, including patient motion, to estimate a PTV margin [6]. Fundamentally, both of these population-based approaches are designed to estimate a margin for the majority of patients, but not a specific patient.

By considering the differences in motion between patients, the PTV margin could be expanded or reduced depending of individual patient requirements. In particular, there may be patient-specific factors (PSFs) that are related to patient motion which can be used to anticipate patients that do not fit the standard population-based PTV margin. A patient may actually require a larger PTV margin, for example, or a smaller one based on their PSF profile. In either case, knowledge of these inter-patient motion variations and their corresponding predictors are important for treatment quality improvement.

1.3 Thesis Scope

The purpose of this thesis is to investigate the feasibility of personalized PTV margins for external beam radiation therapy of the prostate, with particular emphasis on regression-based predictive algorithms [8]. This includes the analysis of prostate motions and rotations, as well as their associated potential predictors, the PSFs. Population-based statistics are calculated to estimate the typical range of motions/rotations as well as investigate the relationship between PSFs and motions/rotations for prostate cancer patients in the BCCA-VIC.

Predictive algorithms are benchmarked for application in predicting patient intra-fraction motion. Clinically relevant motion data and PSFs are generated and used to evaluate the performance of several predictive algorithms, with additional emphasis on patient data requirements and the influence of input data noise. Benchmarked algorithms were applied to actual patient data and evaluated for use in predicting intra-fraction motion along all three spatial directions. The motion predictions generated by the algorithm with the best performance were incorporated, along with other uncertainties in radiation therapy planning, set-up, and delivery, to produce a personalized PTV margin.

Chapter 2 provides an overview of basic radiation therapy. Chapter 3 gives a general overview on statistics and other relevant concepts. Chapter 4 outlines the methodology used in this thesis. Chapter 5 discusses the results of the PSFs and other population-based inferences, the results of the predictive algorithm benchmarking, the results of the predictive algorithms when applied to actual patient data, and the formulation of a personalized PTV margin. Chapter 6 discusses these results, their relevance to personalized prostate EBRT, as well as options for future work. Chapter 7 concludes the thesis and offers thoughts on future work.

2 Radiation Therapy Background

This chapter will cover the necessary radiation therapy background for this research project.

2.1 External Beam Radiation Therapy

External beam radiation therapy (EBRT) describes a collection of treatment methods used to deliver ionizing radiation from outside the patient (as opposed to radiation originating from sources placed inside the patient) in order to kill cancer cells. EBRT includes kV x-ray treatment units (superficial and orthovoltage units) and teletherapy (gamma ray source machines), but is most commonly performed using a medical linear accelerator (linac) [9].

2.1.1 Linac Photon Generation and Modulation

Linacs generate high-energy (MV), ionizing radiation which can be used to irradiate and destroy cancer cells. Ionizing radiation produced by linacs include either photons or electrons, however only photon generation will be discussed here as they are used in the treatment of prostate cancer described in this thesis. Figure 1 shows a diagram of the linac along with its several general components.

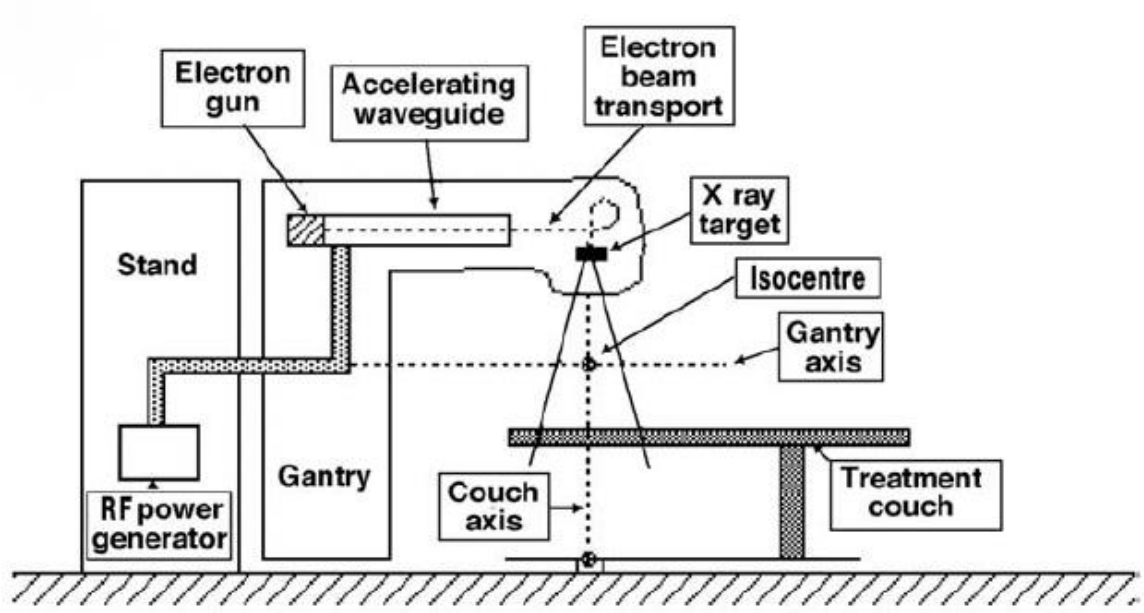

Figure 1: Linac with important components (from Podgorsak, 2005, pg. 140)

Electrons are injected into an accelerating waveguide by an electron gun via thermionic emission. Simultaneous with this injection, a radio frequency (RF) electromagnetic wave, produced by an RF power source, enters the accelerating waveguide. The RF wave accelerates the injected electrons to MeV kinetic energies, where bending magnetic then re-directs them onto a high atomic number target. The deceleration of the electrons in the x-ray target produce high energy bremsstrahlung x-ray radiation, which is further modulated in the linac treatment head for clinical use [9].

Clinical x-ray beam modulation is performed using a number of different components within the treatment head. A flattening filter is typically used in combination with the x-ray target to produce a uniform intensity x-ray beam. Primary and secondary collimators are used to define the maximum extent of the treatment field. A multi-leaf collimator (MLC) is often used as a collimator for further field modulation by defining custom,

dynamic radiation fields for patient treatments [10]. Intensity Modulated Radiation Therapy (IMRT), for example, is a common treatment option where the MLC modulates the intensity of radiation in a given field treatment field to achieve a superior overall dose distribution [11].

2.1.2 Linac Treatment Couch and Coordinate System

The linac is associated with a patient support assembly, often called the treatment couch. The treatment couch typically allows motion along three spatial dimensions (vertical, longitudinal, and lateral). Figure 2 shows a treatment couch along with an associated 3D couch motion coordinate system (Isocentric standard representation, Varian IEC scale). The vertical axis is parallel to the patient's anterior-posterior axis, with the posterior direction defined as positive. The longitudinal axis is parallel to the patient's superior-inferior axis, with the superior direction defined as positive. The lateral axis is parallel to the patient's right-left axis, with the left direction defined as positive.

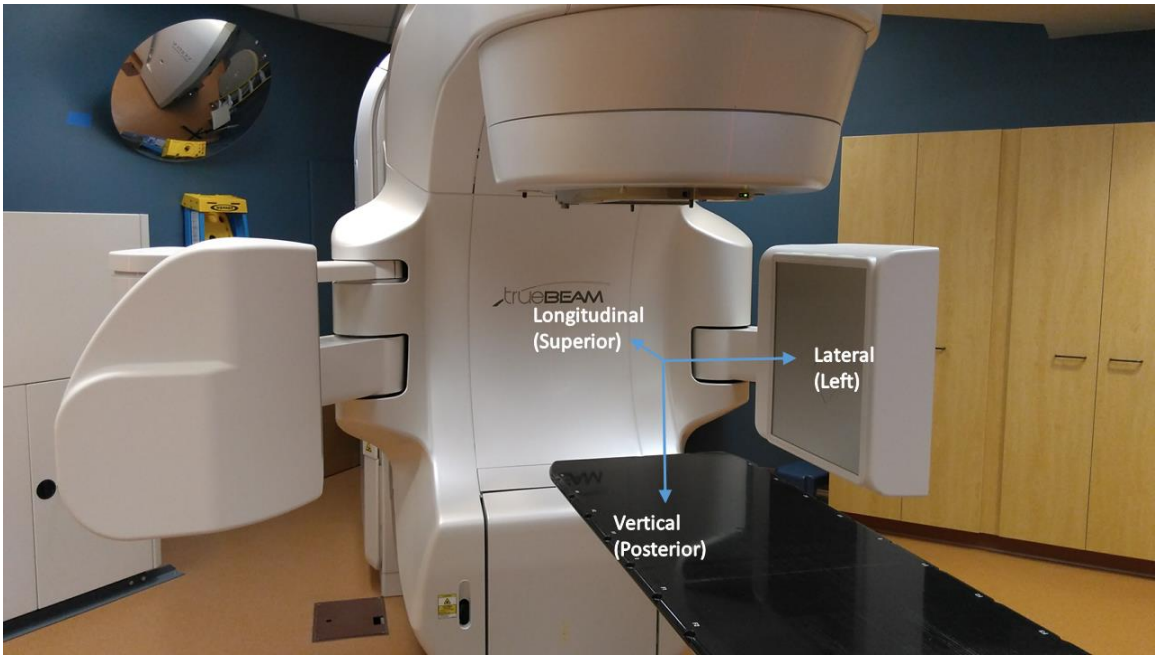

Figure 2: Linac with accompanying coordinate system axes.

While 3D couch motion is more commonly used for routine treatments such as prostate EBRT, 6D couch motion is also possible on more specialized couches. 6D couch motion includes the three motion axes as well as three rotational axes. Rotation about the lateral axis is called the pitch. Rotation about the longitudinal axis is called the roll. Rotation about the vertical axis is called the yaw.

2.1.3 Linac on-board imaging

Modern linacs are equipped with an on-board kV imaging system (OBI) to assist with validation of patient set-up. OBI consists of a kV x-ray source and detector which are both mounted to the gantry opposite each other (Figure 3).

Figure 3: Linac OBI with kV x-ray source and detector labelled.

There are two principle methods of imaging the patient using OBI, which are both used for prostate image guidance. The first is to collect two perpendicular patient images, one along the vertical axis and one along the lateral axis. This method is referred to a perpendicular-

paired or orthogonal-paired imaging and is typically used in conjunction with high density implanted fiducial markers to generate sufficient contrast. A kV image typically delivers 1 – 3 mGy of dose to the patient per image [12]. The second OBI method is a cone beam CT (CBCT). CBCT is performed by continuous x-ray fluorescence and gantry rotation around the patient, collecting several projection images of the patient's internal anatomy. The collected projections are then reconstructed to create a volumetric representation of the patient's internal anatomy. A CBCT typically delivers a dose of 16 mGy to the center of the body and 23 mGy to the body's surface [12].

2.2 External Beam Radiation Therapy Process

The EBRT process involves multiple steps, requiring coordination between physicians, physicists, dosimetrists, and therapists. Patient imaging is first performed to obtain a digital representation of the patient's internal anatomy. This is typically done using a computed tomography (CT) scanner, although other modalities may be used in addition to CT depending on the particular treatment [10].

A physician uses the image dataset collected by the CT scanner to identify and contour the region to be treated (CTV) as well as important organs at risk (OAR). Once the important volumes are defined, a dose is prescribed to the PTV and dose limits are set for the OAR. The CT image dataset is also often used (always used if no physician is present at CT scan) to determine the resultant couch shift between the patient's CT simulation position and the treatment isocenter position (often located within the PTV) [10].

Beam arrangement and parameters are determined by a treatment planning system used by dosimetrists, and sometimes physicists, in a process called plan optimization. Typically, an initial dose distribution to the PTV and surrounding normal tissues is calculated and assessed based on the prescription dose and dose limits. If dose distribution improvements are required, the importance of the different dose constraints are often adjusted to modify the beam parameters, and improve the dose distribution. At VIC, a plan requires approval from a physician and two physicists (including a second check) before it is used for patient treatment.

An approved plan is then transferred to the linac for delivery to the patient. A course of radiation therapy is generally a multi-week process, where delivery of the total prescribed dose to the PTV is divided into the number of fractions. Every fraction the patient is set-up by laser alignment, which is then verified and corrected using OBI, to ensure the accuracy of the treatment.

2.2.1 External Beam Radiation Therapy of the Prostate

EBRT of the prostate at the VICC is typically planned using 5 intensity-modulated treatment fields, although 7 fields may be used for larger patients. The five treatment fields are planned at gantry angles of 100°, 50°, 310°, 260°, and 0°. PTV dose prescription for radiation therapy of the prostate is typically 74 to 78 Gy over 37 to 39 fractions, which equates to 2 Gy per fraction. Figure 4 shows a standard 5-field prostate plan. Figure 5 shows the various motion and rotation directions available to the prostate.

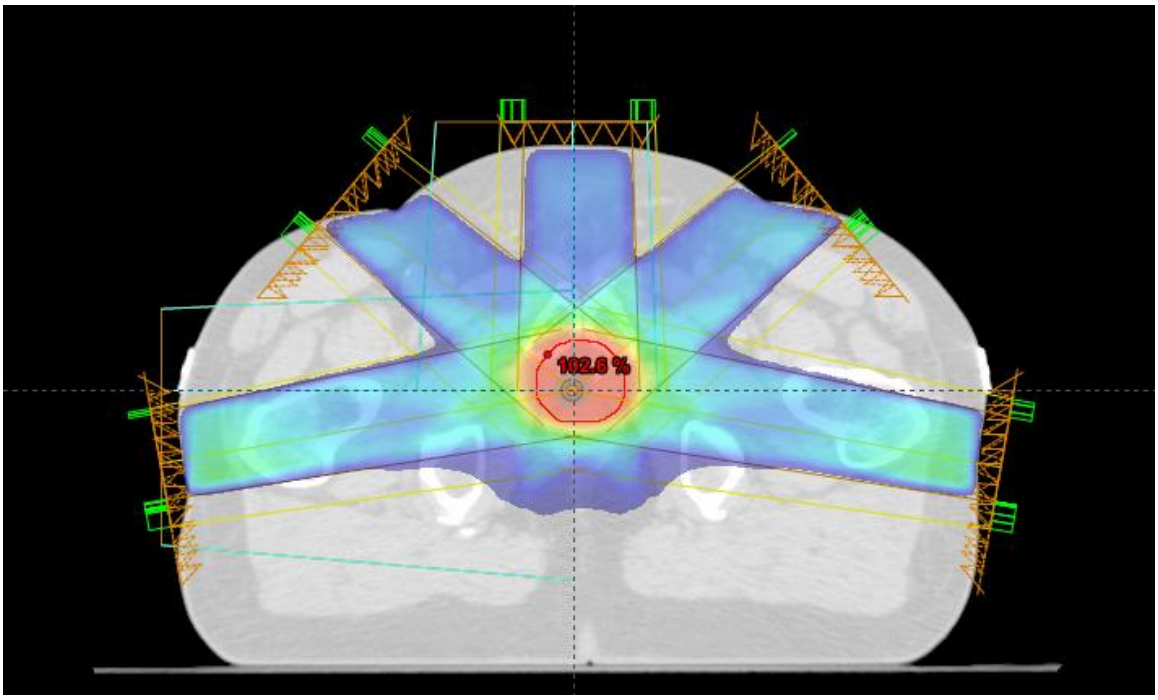

Figure 4: Example of five-field prostate EBRT plan. Five intensity-modulated radiation fields are delivered at gantry angles of 100°, 50°, 310°, 260°, and 0°. The red circle is the PTV contour.

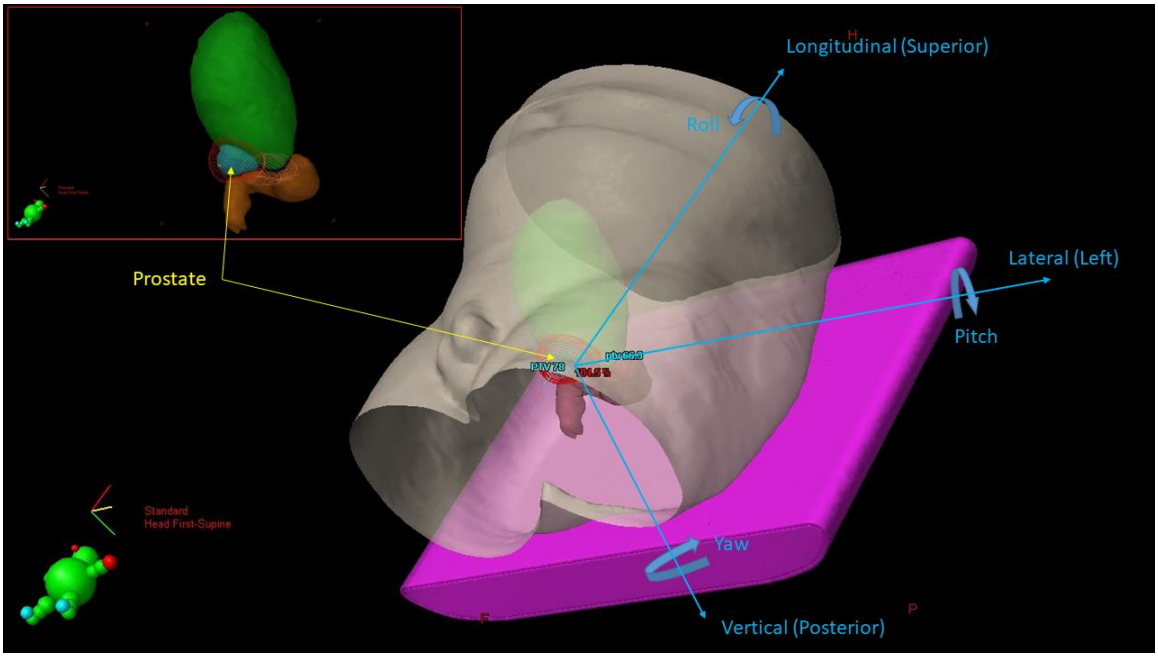

Figure 5: 3D representation of prostate and neighbouring organs, along with motion and rotation axes.

2.3 Volumes and Margins in Radiation Therapy

Figure 6 shows the various volumes and margins defined in radiation therapy as defined in ICRU Report 62.

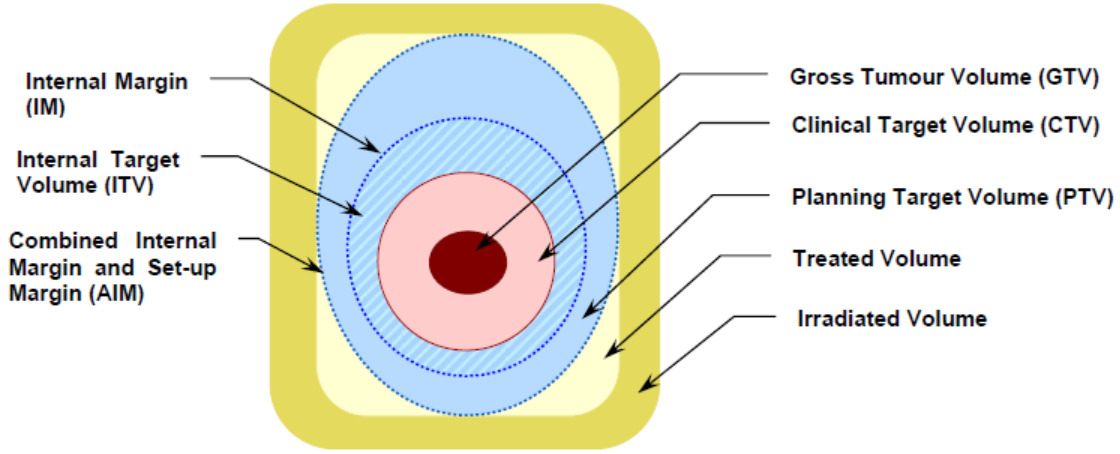

Figure 6: Margins defined in radiation therapy in accordance with ICRU 62 (courtesy I Spadinger).

The gross tumour volume (GTV) is defined in ICRU 62 as "*the gross demonstrable extent and location of the malignant growth."*

The clinical target volume (CTV) is defined in ICRU 62 as "*a tissue volume that contains a demonstrable GTV and/or subclinical malignant disease that must be eliminated. This volume must be treated adequately in order to achieve the aim of radical therapy."*

The planning target volume (PTV) is defined in ICRU 62 as *"a geometrical concept used for treatment planning, and it is defined to select appropriate beam sizes and beam arrangements, to ensure that the prescribed dose is actually delivered to the CTV."*

The PTV therefore includes the CTV plus an additional margin (the PTV margin) to account for uncertainties in treatment set-up and delivery. Conceptually, the PTV margin includes two components: the internal margin and the set-up margin. The internal margin accounts for uncertainties caused by *"physiologic movements and variations in size, shape, and position of the CTV"*. These uncertainties cannot be easily controlled and may include patient breathing, gut peristalsis, bowel movements, and more. The set-up margin accounts for uncertainties caused by patient positioning and beam alignment, which may include patient positioning variation, equipment mechanical uncertainties, and set-up errors in transfer from simulator to treatment unit for example.

The treated volume is the volume that receives a clinically high dose. This includes the PTV and additional tissue due to limitations in radiation therapy technology (e.g. MLC leaf thickness/resolution). The irradiated volume is the additional volume around the treated volume which receives a dose that may be significant to surrounding normal tissues.

2.4 PTV Margin Estimation

2.4.1 Standard Deviation Approach

One method to estimate a PTV margin is to record and monitor the motion of a large number of previous patients in order to estimate the motion of future patients. The mean and standard deviations provide an indication of typical motion to be expected during the treatment. From this the PTV margin is set to account for the motion of a large proportion

of the population. Ideally the only patients not covered by the PTV margin are those who experience some unexpected contributors to motion during their treatment.

2.4.2 van Herk Formula

The van Herk formula is a population-based approach to estimate a PTV margin. It combines all systematic and random errors during radiation therapy planning and delivery to generate a margin that ensures 90% of the patient population receives a minimum dose of 95% to the CTV. [6]

In radiation therapy, a systematic error can be defined as an error that affects all fractions in a similar magnitude and direction throughout the whole course. The result of a systematic error is a difference in the position between the planned and delivered radiation. A random error affects an individual fraction. The result of a random error over the whole course of radiation is an increased spread in the dose distribution. These systematic and random error contributions can be broken down further into several sub-groups: Errors due to target delineation, errors due to set-up, errors due to inter-fraction motion, and errors due to intra-fraction motion.

Target delineation is a purely systematic error since it is identified once at the beginning and is based on inter-physician and intra-physician contouring uncertainty as well as imaging modality resolution.

Set-up errors can be both systematic and random. A systematic set-up error may be caused by fiducial marker drift, patient weight loss, or CTV shrinkage for example. A random set-

up error may be caused by therapist matching uncertainty or by variability in organ (bladder, rectum, etc) fullness for example.

Inter-fraction motion error is the error introduced between fractions due to patient repositioning. This error is effectively eliminated using image guidance.

Intra-fraction error is the error introduced due to target motion after the pre-treatment imaging and prior to, or during the treatment. Intra-fraction errors are typically random errors caused by variability in patient motion. It may also have a systematic component caused by organ drift, however.

The van Herk PTV margin is estimated using the individual contributions of these errors:

$$PTV\ Margin = 2.5\sum_{pop} + 0.7\sigma_{pop}$$

The total population-based systematic error contribution, Σ_{pop}, is calculated by adding the individual systematic error contributions in quadrature:

$$\Sigma_{\text{pop}}^2 = \Sigma_m^2 + \Sigma_s^2 + \Sigma_d^2$$

Σ_s is the systematic error due to set-up uncertainty.

Σ_d is the systematic error due to target delineation.

Σ_m is the systematic error due to organ motion, which is estimated as the standard deviation of all individual patient mean intra-fraction motions (an indication of the reproducibility of mean intra-fraction motion).

The total population-based random error contribution, σ_{pop}, is calculated:

$$\sigma_{pop}^2 = \sigma_m^2 + \sigma_s^2$$

σ_m is the random error due to organ motion, which is estimated as the root-mean-square of all individual patient intra-fraction motion standard deviations (an indication of the average random motion an individual patient experiences).

σ_s is the random error due to set-up uncertainty.

3 Statistical Analysis Background

This chapter will cover the basis of the data analysis and techniques used in this work. This includes an introduction to data preprocessing and basic statistics. An introduction to predictive algorithms used for regression is also included with focus on model selection, as well as training and cross-validation.

3.1 Overview

Data analysis is the general process of investigating data to extract useful information. This is typically performed by following a general methodology: Data preprocessing, information extraction, information evaluation, and information presentation. This section of the chapter introduces important elements and definitions relevant to the data analysis process in this work.

3.2 Data Preprocessing

Data pre-processing is the broad process through which data is prepared to be investigated for useful knowledge. This is an important step in the data analysis process and often involves some or all of data cleaning, data integration, and data transformation [13]. A brief introduction to these concepts is provided in this section. For specific approaches used in this work, please refer to the methodology section.

Data cleaning deals with common data collection problems such as missing data, inconsistent data, and noisy data. There are many methods to deal with these problems and the approach is ultimately chosen based on the problems specific to the task. Missing and otherwise inconsistent data can be dealt with, for example, by ignoring the whole patient's data entry, filling the patient's entries manually, filling the missing entry with a summarizing statistic (mean, median, etc.), or filling the missing entries with some dummy constant (such as not-a-number). Noisy data (data entries with a large variance), can be cleaned by binning or otherwise aggregating the data.

Data integration is the process through which data from multiple sources (and likely data structures) are bought together. The two most common difficulties in this process are properly associating two or more datasets, and merging different data structures.

Data transformation is the process through which the raw input data is restructured in a way more appropriate for extracting useful information. Data transformation techniques may include summarizing input data (mean, min, max), creating new fields more relevant to the study from pre-existing fields (for instance by subtraction), and data scaling.

3.3 Basic Statistics Definitions

Sample Mean

For a set of n observations, the sample mean, $\bar{x}$, of the set of x_i is defined:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Sample Standard Deviation

For a set of n observations, the sample standard deviation, σ, of the set of x_i is defined as:

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Where $\bar{x}$ is the sample mean of the set of x_i.

Pearson Correlation

The Pearson correlation, r, tests the degree to which two variables, x and y, vary linearly with each other:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

A perfect positive correlation yields a result of +1, a perfect negative correlation yields a result of -1, and correlation of 0 indicates no relationship between the variables.

Spearman Correlation

The Spearman correlation, ρ, tests the degree to which the rankings of two variables, x and y, vary linearly with each other:

$$\rho = \frac{\sum_{i=1}^{n}(rank(x_i) - rank(\bar{x}))(rank(y_i) - rank(\bar{y}))}{\sqrt{\sum_{i=1}^{n}(rank(x_i) - rank(\bar{x}))^2}\sqrt{\sum_{i=1}^{n}(rank(y_i) - rank(\bar{y}))^2}}$$

Where the rank of a variable refers to the ordinal position of its values, rather than their numerical values. For example:

$$x = \{1, 2, 5, 20, 18, 36\}$$

$$rank(x) = \{1, 2, 3, 5, 4, 6\}$$

Similar to the Pearson correlation, a perfect positive rank correlation yields a result of +1, a perfect negative rank correlation yields a result of -1, and a correlation of 0 indicates no relationship between the rankings of the variables. A Spearman correlation is useful here as it does not assume an underlying linear relationship between the variables. It is also more robust to outliers (extreme values) as it only uses the rankings of the collected samples.

3.4 Predictive Algorithms

3.4.1 General Overview

Predictive algorithms (or predictive models) are models that use pre-existing data to make predictions about new data. In reality, there are several names for tools that perform these

tasks ranging from machine learning, statistical learning, predictive analytics, and several more. The phrase predictive algorithms will be used for this thesis.

Defined this way, predictive algorithms are a type of supervised learning, which means that they require input data consisting of both a set of input variables and an associated output variable (or labelled data). This is in contrast to unsupervised learning algorithms which require only the set of input variables, but no associated output variable (unlabelled data).

Supervised learning algorithms are further specified based on the type of output variable they predict. Two more commonly predicted outputs are categories and numbers. Algorithms that predict a category are called classification algorithms, whereas algorithms that predict a number are called regression algorithms. The focus of predictive algorithms in this thesis are regression algorithms.

3.4.2 The Predictive Algorithm Process

Feature Selection

Feature selection (along with other data preprocessing) is an important step in the predictive algorithm process. This step involves choosing the input variables to be used as predictors for the models. One common approach to feature selection is to evaluate candidate input variables based on their correlation with the output variable of interest, where only the input variables above some threshold minimum correlation are then included as predictors for the model. Model performance is often best when only the most strongly correlated predictors are used as input to the model.

Scaling of Input Data

Scaling of input data is critically important to the performance of some predictive algorithms and is often required to ensure that different input features have equal importance to the output despite potentially large differences in range between inputs. Typically, each input feature is scaled by removing the mean and dividing by the standard deviation:

$$x_i^{scaled} = \frac{\left(x_i^{unscaled} - \bar{x}_i\right)}{\sigma_{x_i}}$$

Where $\bar{x}_i$ is the mean value and σ_{x_i} is the standard deviation of a particular input feature, x_i. This makes each input feature roughly Gaussian in appearance.

Model Selection

Initial model selection is done based on the type of data (labelled or unlabelled) as well as the problem to be solved (regression, classification, etc.). A large number of models satisfy these criteria however, and it is useful to narrow down the list further in order to identify a few candidate algorithms more appropriate for our problem.

When considering the feasibility of an algorithm for a particular problem, there are a few common input data considerations: The number of data points available (or number of samples), the number of variables for each sample (or input features), and the degree of noise in the data. The number of samples and features, in particular, are often known. The amount of noise in the input data may be more challenging to estimate.

These data descriptors can be used to gauge a suitable model for a given problem based on the model's complexity. Model complexity is essentially the number of degrees of freedom available to the model. A more complex model has a larger number of degrees of freedom, whereas a less complex model has a smaller number of degrees of freedom. The optimal model is one that has enough complexity to capture the underlying signal in the data, but not so much complexity that it makes predictions based on noise in the data. A complex algorithm is therefore more likely suitable to a dataset consisting of a large number of samples and/or a small number of features, whereas a less complex algorithm is more likely suitable to a dataset consisting of a small number of samples and/or a large number of features.

Model complexity and selection is intimately related to a problem called the "Bias-Variance Trade-Off" [8], which essentially stems from the fact that the input data is stochastic and therefore algorithm training is done in a way that is not necessarily generalizable to future data. While this is a general concept for any loss function [14], there is a well-known prediction error decomposition for a mean square loss function:

$$Prediction\ Error = Irreducible\ Error + \underbrace{Bias^2 + Variance}_{Reducible\ Error}$$

Bias is the prediction error introduced due to the mean prediction value of the model being different from the ideal prediction value. Variance is the error introduced due to the used prediction value being different from the mean prediction value. Irreducible error is the

error introduced due to the fundamental randomness in the input data. More specifically, given a particular set of inputs x associated with an output y, an ideal model f(x) which predicts the output y imperfectly due to some intrinsic noise, and an estimate of the ideal model $\hat{f}(x)$, the bias, variance, and irreducible error can be broken down as follows:

$$y = f(x) + \epsilon$$

$$Bias = E\left[\hat{f}(x)\right] - f(x)$$

$$Variance = E\left[\hat{f}(x) - E\left[\hat{f}(x)\right]\right]^2$$

$$Irreducible\ Error = Variance(\epsilon)$$

Generally, a model with more complexity (degrees of freedom) has a lower bias error, but a higher variance error. It is able to capture more accurate information on average, but is subject to sampling variability. Conversely, a model with less complexity generally has a higher bias error, but a lower variance error. It captures less information on average, but is also more robust to sampling variability.

Training and Cross-Validation

As explained in Hastie et al. [8] , training an algorithm and validating its performance is typically done by dividing the whole available dataset into either two or three subsets. Three data sets are ideal: A training dataset to select/update the algorithm parameters, a validation dataset to calculate prediction error for model selection/parameter tuning, and test dataset as a final check that the model performs well. However, the data is often divided into only two datasets, the training dataset and the validation dataset when data is sparse.

Evaluating algorithm performance on an independent dataset is important because algorithm performance can become over-estimated if it is validated on the same dataset it was trained on. This provides a more accurate indication of the algorithm's generalizability to making predictions on future data.

Leave-One-Out Cross Validation (LOO-CV) [8] is type of cross-validation that is important in this thesis work. This is a commonly used technique to evaluate algorithm performance when data is limited. It uses all but one data point for training the algorithm. The single remaining data point is used to validate the algorithm's performance. This process is repeated for all permutations of training and validation data splits, providing an exhaustive measure of the algorithm's performance.

LOO-CV estimate of prediction error for a model f, is calculated as the mean value of the validation error, L, on each of the n individual permutations:

$$CV(f) = \frac{1}{n}\sum_{i=1}^{n} L\left(y_i, f^*(x_i)\right)$$

Where f^* is the model f trained without the i^{th} training data point. The validation error, L, on each of the individual permutations is usually calculated using either the mean absolute error (MAE) or the mean squared error (MSE):

$$MAE: L\left(y_i, f^*(x_i)\right) = |y_i - f^*(x_i)|$$

$$MSE: L\left(y_i, f^*(x_i)\right) = \left(y_i - f^*(x_i)\right)^2$$

Parameter Tuning

Several of the algorithms used in this thesis are adaptable through modification of different tuneable parameters called hyperparameters – parameters used in an algorithm, but external to the data itself. Adjustment of these hyperparameters often modify the goal of algorithm training in order to improve the generalizability of the algorithm to future data. In this sense, parameter tuning is closely related to the Bias-Variance Trade-Off discussed previously.

Parameter tuning is often performed using cross-validation as discussed previously, where the average validation error of an algorithm is estimated for different hyperparameter values. The optimal choice of hyperparameter is the one which yields the lowest average validation error, as this is an indicator of the generalizability of the parameter choice towards predictions on future collected data.

4 Methodology

The methodology for this work was divided into four sections:

1. The collection of patient intra-fraction motion data coupled with patient demographic/treatment data;

2. The pre-processing and analysis of this data;

3. The investigation into candidate predictive algorithms suitable for predicting a patient's intra-fraction motion based on their patient-specific factors (PSFs);

4. The generation of a personalized PTV margin.

4.1 Data Collection

Demographic data, motion data, and other relevant treatment data was collected for 21 prostate cancer patients to be treated with 74Gy to 78Gy over either standard (37-39 fractions) or hypo-fractionated (26-28 fractions) radiation therapy schedules. For organizational purposes, the collected data was broken down into two subgroups, time-independent data and time-dependent data. Time-independent data was defined as data that remains constant for a patient over an entire course of radiation therapy, whereas time-dependent data was defined as data that varies on a fraction-to-fraction basis for each patient.

4.1.1 Time Independent Data

Time-independent data was defined as data that remains constant over an entire course of radiation therapy. This included data relevant to clinical decision making (clinical data), demographic data, and other data that does not change on a fraction-to-fraction basis (other time-independent data).

Clinical data included factors commonly used for staging prostate cancer, or recorded during the staging of prostate cancer. These clinical factors were PSA score, primary Gleason score, secondary Gleason score, total Gleason score, number of positive cores, total number of cores sampled, and cancer staging.

Demographic data included any patient descriptors that could be collected during physician consults and deemed relevant to this study. The demographic factors were age, weight, height, body mass index (BMI), ECOG status, and whether the patient has any of diabetes, irritable bowel syndrome (IBS), chronic obstructive pulmonary disease (COPD), or implanted prosthetics (implants).

Other time-independent data include any other relevant factors that did not change over the course of radiation therapy. There was only one extra factor recorded which was the number of implanted fiducial markers.

A brief overview of all of these factors is included below along with any associated challenges in collection.

PSA Score

PSA Score is short for Prostate Specific Antigen Score and is a measure of the concentration [ug/L] of a particular protein (antigen) produced by the prostate. PSA score is available for all prostate cancer patients and is measured through a blood test [15]. Scores typically range between 0 ug/L and around 10 ug/L, however in rare cases they can be much higher. [16] It is known that a higher concentration of this antigen or *higher PSA score* is indicative of an increased likelihood of prostate cancer.

One of the challenges in using PSA score for this work is that (in order to assess the cancerous activity in the prostate) a patient often has many PSA scores taken leading up to their treatment. This makes it difficult to choose an identical score across all patients. This is further complicated by the fact that a number of patients undergo surgery and/or hormone therapy prior to receiving radiation therapy, which can decrease the PSA score dramatically. With these issues in mind, a decision was made in consultation with a physician to choose the most recent PSA score prior to the start of the course of radiation therapy. Since the information we are searching for is intra-fraction motion, it is the score most likely to be correlated with motion.

Gleason Score

Gleason score is a measure of the pathology of the prostate tissue. Two scores are typically assigned by a pathologist following prostate biopsy, a primary Gleason score and a secondary Gleason score. Primary Gleason score is the score assigned to the most common class of prostate cancer cells seen in biopsied tissue. Secondary Gleason score is the score

assigned to the second most common class of prostate cancer cells. Both primary and secondary Gleason scores range from 1 to 5, with 1 being the least pathological and 5 being the most pathological. The sum of these two Gleason scores provides another metric called the total Gleason score. [17]

Number of Cores Biopsied and Number of Positive Cores

The number of cores biopsied is the number of prostate tissue sample biopsies taken for the Gleason score analysis. Typically, 10-12 biopsy samples are taken, although there is variation to optimize cancer detection sensitivity vs biopsy side-effects [18] . The number of positive cores is the number of biopsied cores that return a positive test result for cancer.

Cancer Staging

Cancer staging is the stage assigned to the patient's prostate cancer. Staging is done by a physician and classifies the extent of the cancer. Prostate cancer staging is typically done with three qualities in mind: The extent of the tumour (T) volume, the involvement of the lymph nodes (N), and whether the cancer is metastatic (M). This is called the TNM staging. [19]

While staging was recorded for each patient, it was not included in the work due to the specificity of the staging coupled with the low number of patients. In particular, there were 11 unique stages in the group of 21 patients.

Age

The age of each patient was recorded. Prostate cancer is much more common in older men. Men over 50 years old are at highest risk with most men being diagnosed over age 65. [20]

Height, Weight, and BMI

The height and weight were recorded for all patients. The average adult male in Canada has a height and weight of around 175cm and 84kg, respectively. [21] The height and weight values were also used to calculate the BMI using the accepted definition [22]:

$$BMI = weight/height^2$$

BMI was a factor of particular interest in the study as previous research has suggested a correlation between it and prostate intra-fraction motion [23].

Number of Fiducials

This is the number of fiducial markers implanted into the prostate. Generally the recommended minimum number of seeds is 3, although more may be inserted if fiducial loss or migration is suspected [24].

ECOG Status

ECOG status is an all-encompassing measure of the impact of disease (and treatment) on the daily life of a patient. The scale ranges from zero to five, where a score of zero means the patient experiences no impact on their daily life and a score of five means patient is

dead. In between scores provide indication into how much assistance patients require to perform routine tasks [25].

Diabetes

Diabetes is a condition characterized by high blood glucose levels due to inadequate insulin production by the pancreas. Permanent forms of diabetes are classified as either *Type 1,* where the patient's insulin production has been totally ceased due to damage from their own immune system, or *Type 2,* where inadequate amounts of insulin are produced or used due to other reasons. The majority of cases are type 2, representing about 85% to 95% of cases in developed countries. At the biochemical level, unregulated changes in blood glucose levels cause changes in the amount of water drawn into and out of cells. It is reasonable to consider that this swelling or shrinkage of tissue may affect the motion of the prostate to some degree. [26,27]

Diabetes was treated as a Boolean quantity. Presence of any diabetes, whether type 1 or type 2 was considered a 'yes' for diabetes.

IBS

IBS is the most commonly diagnosed gastrointestinal condition, affecting 9% - 23% of the world's population. It is characterized by abdominal pain, straining, urgency, and bloating, all of which may relate to change in typical prostate motion. IBS may also be associated with changes in peristalsis – the involuntary gut contractions that move food through the digestive track – which may affect prostate motion. [28,29,30]

IBS was treated as a Boolean. Presence of IBS was considered a ‘yes’ for IBS.

COPD

Chronic Obstructive Pulmonary Disease (COPD) is a common disease accounting for approximately 5% of deaths worldwide in 2015. It is characterized by airflow limitations to the lungs and presents with symptoms such as laboured breathing and chronic cough [31]. These symptoms may be associated with increased prostate motion during treatment due to patient discomfort and/or motion.

COPD was treated as a Boolean. Presence of COPD was considered a ‘yes’ for COPD.

Implants

Implants refer to whether a patient had prosthetic replacements. This was recorded as a Boolean, however zero patients ultimately had implants so this has not been consequential.

Race

Race was initially planned for inclusion in the study, as there are known differences in prostate cancer incidence between race [32], but ultimately wasn’t recorded for any of the patients involved and was therefore not used.

4.1.2 Time Dependent Data

Time dependent data includes motion data collected at each fraction, along with data about the presence of rectal gas, and fraction timestamp information.

Prostate motion data was collected through the detection of gold fiducial marker motion. Each patient had 3 gold fiducial markers inserted into their prostate prior to their treatment planning to act as a surrogate for prostate motion. These fiducial markers have a higher density and higher atomic number compared to tissue and, as a result, are easily detectable in tissue using kV imaging. Patients were imaged using either orthogonal paired kV images or cone beam CT.

The imaging was performed immediately prior to treatment and again immediately after treatment. Prior to each fraction, imaging was used to position the patient's prostate relative to the prostate position at time of CT simulation by fiducial marker matching. Fiducial marker matching is an automated process which detects the high contrast fiducial markers in the images and aligns them with their position in the digitally reconstructed radiographs (DRR) produced from the simulation CT. This provides a resultant necessary couch shift.

The couch shift was calculated using two methods. The first method was to align the center-of-mass (COM) position of the fiducial markers on the day of treatment with the simulation CT COM position, such that the relative distance between the fiducial markers is unchanged. The resultant couch shift was therefore calculated as the difference between the two COM positions. The second method was to match the position of each individual

fiducial marker independently. The resultant couch shift was then calculated as the average difference between the positons of the three fiducial markers. This method was used if fiducial drift was suspected, causing the relative position between the fiducials to change.

Ultimately, a couch shift was calculated for both the 3D case and the 6D case, and both were recorded as an indicator of inter-fraction motion. However, only the 3D couch shift was used to actually shift the couch for all the patients. The result of the 3D couch shift was that inter-fraction motion of the prostate was effectively eliminated, however inter-fraction rotational errors remained.

After each fraction, imaging was used to again match the patient's prostate to the treatment planning position of the couch. Couch shifts were calculated for both the 3D case and the 6D case, and again both were recorded. The difference between the post-treatment position of the prostate and the pre-treatment position of the fiducial markers is an indicator of intra-fraction motion that occurred during treatment.

In addition to motion information, the presence of gas prior to treatment and after treatment was recorded every fraction as well as a number of time stamps during each fraction. Presence of gas was recorded as a Boolean quantity; the patient either had or did not have gas. This was evaluated through visual inspection of the pre-treatment and post-treatment images, indicated by air gaps in the rectum area.

Time stamps were also taken each fraction at the time of first beam, at the time of the last beam, and the total time the patient's file was open. These times were recorded to investigate the relationship between prostate motion and treatment times.

4.2 Data Analysis

Broadly, data analysis is the process through which one searches for knowledge by investigating a collection of data. This process involves many steps, but can be grouped into three more general steps: Preprocessing of the data, extracting knowledge from the data, and evaluating and presenting the findings. All of these steps will be discussed in the following sections, but firstly, a tool to perform these tasks had to be chosen.

4.2.1 Overview of the software environment

The software used for data analysis purposes was the Pandas module within Python [https://www.python.org/]. Python is a free, open source programming language, and Pandas is a data analysis library available for use within the Python environment [http://pandas.pydata.org/]. It is built using several tools from pre-existing Python libraries, most notably Numpy [http://www.numpy.org/], the standard numerical library in Python, and Scipy [https://www.scipy.org/], the standard scientific library in Python.

4.2.2 Data Preprocessing

Data preprocessing is the first – and arguably the most important – step in the data analysis process as it lays the foundation for the rest of the analysis. This process can be thought of as maximizing the signal-to-noise within the collected data. It addresses missing or otherwise anomalous or inconsistent data, pooling data from multiple sources, choosing

which elements of the collected data are most important, and structuring the data in a more useful way to extract information. Data was preprocessed using all of these mentioned elements in at least some capacity.

The time-independent and time-dependent data was the input data to the pre-processing phase. The first task was to organize this data in a manner so that it could be read-in and analyzed consistently for all patients. This was done for both the demographic and motion data files and was completed by creating .csv file templates for the motion and demographic input files (see appendix). The process of taking the two data files, checking for problems such as units or missing information, and putting them into a template csv file takes roughly 2-3 min per patient, depending on the non-uniformity of a given patient's data.

Once data was imported into the program, additional modifications were required. Patient files with missing data entries ("n/a") were converted to nan ("not-a-number") and were consequently omitted from statistical analysis. Patient files with inconsistent data entries (e.g. "no", "N", "n", etc.) were made to be homogeneous. Data-type problems (e.g. floating point vs. string) caused by the presence of missing or inconsistent data were also corrected. Any additional result-specific data manipulations are included in the sections below.

4.2.3 Searching for Knowledge

Identification of Patient Specific Factors

The inputted data was investigated on several fronts. The first was to identify if and what PSFs correlated with intra-fraction motion. This analysis is useful to identify important features (PSFs) to use as input to predictive algorithms.

An additional pre-processing step was performed for this section. The intra-fraction motion data (and other time dependent data such as time stamps and gas) from each fraction was collapsed into summary statistics such as the mean, standard deviation, min and max. This data *transformation* was necessary to compensate for the fact that the motion measured by the fiducial markers are only a snapshot into the true motion of the prostate. By looking at those motion snapshots over the course of multiple fractions, one can have more confidence that the fiducial motion is an accurate representation of the actual prostate motion.

Two correlation tests were used for this analysis, a Pearson (r) correlation and a Spearman (ρ) correlation.

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

$$\rho = \frac{\sum_{i=1}^{n}(rank(x_i) - rank(\bar{x}))(rank(y_i) - rank(\bar{y}))}{\sqrt{\sum_{i=1}^{n}(rank(x_i) - rank(\bar{x}))^2}\sqrt{\sum_{i=1}^{n}(rank(y_i) - rank(\bar{y}))^2}}$$

The relationship of interest in this section was between PSFs and any intra-fraction motion, intra-fraction rotation, or inter-fraction rotation. Of particular interest was the relationship between BMI and intra-fraction motion as previous studies suggest a potential relationship between them. [23]

Trends in Time Dependent Data

A second method of searching for information in the data is by looking at the time dependent data or the data recorded on a fraction-to-fraction basis. The time dependent data includes information about fiducial motion and rotations along all directions and axes, as well as time stamps information and the presence of gas. In particular, previous research has shown a relationship between fraction duration and the amount of intra-fraction motion during treatment [33]. This time-dependent analysis is also of special note because it contains a reasonably large sample size, in contrast to what is available in other sections. The 21 patients used for this work account for a combined 678 fractions worth of information.

These 678 fractions were concatenated into one structure, called the ‘population dataset’, and then Pearson and Spearman correlation tests were performed to estimate the degree of correlation between any of the time-dependent parameters.

An additional preprocessing steps was used in this section. A ‘fraction number’ column was added to each patient to investigate trends in the population that occur over a course of radiation therapy.

Rectal Distension

The presence or change in the presence of pre-treatment and/or post-treatment gas was investigated for effect on motion and rotation, as previous work has shown a link between gas and prostate motion [34]. Gas was recorded as a Boolean quantity before and after treatment fraction based on the presence of air gaps in the rectum before and after treatment. The population dataset was grouped based on pre-treatment and post-treatment gas permutation as described in Table 1:

Acronym	Pre-treatment Gas Present	Post-treatment Gas Present
N	*No*	-
Y	*Yes*	-
NN	*No*	*No*
NY	*No*	*Yes*
YN	*Yes*	*No*
YY	*Yes*	*Yes*

Table 1: Gas permutation acronyms.

For each gas permutation, the mean and standard deviation of intra-fraction motion, intra-fraction rotation, and inter-fraction rotation were calculated and compared.

Population-Based Motion Statistics

Population-based statistics were calculated using the population dataset. These include the mean and the standard deviation of intra-fraction motions, intra-fraction rotations, and inter-fraction rotations. The directionality of these motions and rotations were also explored by comparing the mean and the standard deviations of opposite directions along the same axis. For example, along the vertical axis, the prostate may move anteriorly to a

greater extent than it moves posteriorly. The directionality statistics were calculated by pooling the 'positive' direction fractions for a particular motion/rotation and comparing the mean and standard deviation to the 'negative' direction fractions for that motion/rotation.

4.3 Predictive Algorithms for Intra-Fraction Motion

4.3.1 Overview of Software Environment

The software used for this component of the project was the Scikit-learn module [http://scikit-learn.org/stable/] within Python. Scikit-learn is a free machine learning package available to download for use within the Python environment.

4.3.2 Predictive Algorithm Selection

Several regression algorithms were investigated for potential use in predicting a patient's intra-fraction motion based on their PSFs. These algorithms were initially chosen to be tested based on the problem to be solved (regression problem) and the properties of the available (labelled data, small number of samples, small number of features). A brief introduction into the algorithms selected for evaluation is included below. More detailed algorithm descriptions can be found in various textbooks [8,35] .

Linear – The linear model is a standard linear regression algorithm. It predicts an output value f(X) given a set of input values X:

$$f(X) = \beta_0 + X\beta$$

Here β_0 is the intercept, f(X) is the predicted intra-fraction motion along a single direction, and X is the set of input features, which are the PSFs:

$$f(X) = predicted\ intrafraction\ motion$$

$$X = \begin{bmatrix} PSA \\ BMI \\ \dots \\ etc. \end{bmatrix}$$

The coefficients β are the weights associated with each PSF:

$$\beta = \begin{bmatrix} PSA\ Coef \\ BMI\ Coef \\ \dots \\ etc. \end{bmatrix}$$

The coefficients are an indication of the importance of a particular PSF to the predicted intra-fraction motion. These coefficients are calculated by minimizing a residual sum of square (RSS) objective function:

$$RSS(\beta) = \sum_{i}^{N} \left(y_i - f(x_i)\right)^2$$

Here y_i is the true intra-fraction motion for the i^{th} sample and $f(x_i)$ is the predicted intra-fraction motion for the i^{th} sample.

Compared to the other linear algorithms, the linear model is the most sensitive to fine-detail in the input data, but it is also the most susceptible to noise.

Ridge – The ridge model is a regularized linear regression algorithm. Similar to the linear model, it predicts an output value f(X) given a set of input values X:

$$f(X) = \beta_0 + X\beta$$

Where f(X), X, β_0, and β are defined identically as in the linear case. The ridge model differs from the linear model in the way the coefficients are determined. The ridge model has an additional term in the RSS objective function that penalizes the size of the coefficients associated with the input features:

$$RSS(\beta) = \sum_{i}^{N} \left(y_i - f(x_i)\right)^2 + \alpha \|\beta\|_2$$

$$\|\beta\|_2 = \sqrt{PSA^2 + BMI^2 + \ldots etc}$$

Here y_i is again the true intra-fraction motion for the ith sample and $f(x_i)$ is the predicted intra-fraction motion for the ith sample.

The result of the additional term in the objective function is that the coefficients are smaller. As a result, the predictions are therefore more robust to noise because the variance in any

one PSF input has a smaller effect on the predicted motion output. The parameter α controls the trade-off between the two components of the objective function, and is chosen by cross-validation [8]. This is discussed in the background section and later in the discussion section, but by default α = 1.0.

LASSO – The LASSO is a regularized linear regression algorithm nearly identical to the ridge. The only difference between the two is that the LASSO penalizes the size of coefficients based on the sum of absolute values instead of the sum of squares:

$$RSS(\beta) = \sum_{i}^{N} \left(y_i - f(x_i)\right)^2 + \alpha \|\beta\|_1$$

$$\|\beta\|_1 = |PSA| + |BMI| + \ \ldots \ etc$$

The LASSO penalty term is more severe compared to the ridge model. The majority of coefficients are often zero. LASSO has use in feature selection for this reason, as it assigns non-zero coefficients to only the most valuable features. However, it does not deal with co-linearity between inputs well and will often choose one coefficient to be non-zero and set the rest to zero.

The parameter α again controls the trade-off between the two components of the objective function, and is chosen by cross-validation. This is discussed in the background section and later in the discussion section, but by default α = 1.0.

Linear Support Vector Regression – The linear support vector regression (SVR) algorithm attempts to fit a linear model to training data similar to the previous algorithms:

$$f(X) = \beta_0 + X\beta$$

Where f(X), X, β_0, and β are defined identically as in the linear case. The coefficients are determined differently, however, and are based on two conditions:

1. The coefficients are as small as possible (making the function as robust to noise as possible);
2. Training data points that differ from their predicted value by less than some user-specified value ε contribute no cost to the objective function (OF).

This is a constrained optimization problem. The OF is minimized:

$$OF = \frac{1}{2}\,\|\beta\|^2 + C\,\sum_i (\alpha_i + \alpha_i^*)$$

Subject to constraints:

1) $y_i - f(x_i) \leq \epsilon + \alpha_i$
2) $f(x_i) - y_i \leq \epsilon + \alpha_i^*$
3) $\alpha_i, \alpha_i^* \geq 0$

α_i and α_i^* are the cost of using a model $f(X)$ that allows training point deviation greater than the user-specified ϵ tolerance. Here, ϵ was set to be 0.5mm, based on the resolution of a margin assigned to a CTV in the Eclipse treatment planning system. α_i is therefore an intra-fraction motion deviation greater than 0.5mm in the case where the true intra-fraction motion is greater than the predicted intra-fraction motion. Whereas α_i^* is an intra-fraction motion deviation greater than ε in the case where the true intra-fraction motion is less than the predicted intra-fraction motion.

The parameter C determines the relative importance between the function, f(X), being robust and the degree to which training data point deviations greater than ε are allowed. This is chosen with cross-validation as explained in the background section. Mathematical details including the solution to the constrained optimization can be found elsewhere [8,35,36].

k – Nearest Neighbour (kNN)

The kNN regression model is a simple model that works by association. It examines the distance (e.g. Euclidean distance) in parameter space between a test point and its 'k' nearest neighbours. The mean of the output values associated with those k nearest neighbours is used to predict an output value for the test point.

By default, k=5 nearest neighbours was used. A test patient's intra-fraction motion was therefore predicted by determining which k=5 existing patients are most similar to the test patient, based on the PSFs. The mean intra-fraction of the k=5 most similar patients was

then used as the predicted intra-fraction motion for the test patient. Because nearest neighbour evaluation is performed using a distance metric, it is crucial that the input data is scaled to account for different units/scales for each PSF.

Multi-Layer Perceptron (MLP)

The MLP is a regression tool based on neural networks. It receives an input vector of parameters which are fed through a single hidden layer of hidden neurons into an activation function and added linearly to create a prediction for the output. The difference between the output prediction and the true output is then fed back into the network to update the weights in the algorithm using backpropagation. A sigmoid function activation function was used making this identical to an artificial neural network (ANN).

This is a currently popular type of algorithm and is the foundation of 'deep learning'. It was included to provide contrast with the other models used here. As it has many more parameters than the other models, it requires much more data before it begins to work effectively. By default, 1 hidden layer containing 100 hidden neurons was used, which is not an unreasonable starting point [8], especially when the purpose is to explore a model that behaves differently from the others used. In practice, 1 hidden layer is sufficient and the number of hidden neuron are typically selected using cross-validation.

4.3.3 Training and Validation of Algorithm Performance

When training an algorithm, one typically leaves out a number of data points for validation purposes. This is done because algorithm performance becomes over-estimated if it is

validated using the same dataset it was trained on. In this work, the training and validation of algorithm performance was done using Leave-One-Out Cross Validation (LOO-CV), is a commonly used technique when data is limited. For more detail refer to the statistical background section.

The principle performance metric used to report the algorithm's performance was mean absolute error (MAE). This is one of two commonly used metrics; the other is the mean squared error (MSE). The MAE was chosen primarily for ease of error interpretation as its units are physically more meaningful.

4.3.4 Benchmarking of Predictive Algorithms

To estimate the performance of these algorithms on a similar dataset, a model of intra-fraction motion was created. The model simulated the fiducial motion based on clinically relevant PSFs. The PSFs were randomly sampled and then fed into the predictive algorithms. The predicted output motion was then compared to the actual simulated output motion.

Generating Motion Data

Simulated data was generated using a range of PSF values. Those PSFs modulated in the simulation include: PSA score, primary Gleason score, secondary Gleason score, total Gleason score, total cores biopsied, number of positive cores, staging, age, weight, height, BMI, diabetes, IBS, COPD, Implants, and ECOG status. From the algorithms' standpoint, this represents several float inputs, several Boolean inputs, one categorical input (stage),

and two inputs dependent on other input values (total Gleason and BMI). For simplicity, each of the PSFs were randomly sampled from a uniform distribution based on a clinically relevant range. For example, height was sampled between 150 cm and 200 cm, whereas primary Gleason was sampled between 1 and 5, inclusive. The range of parameters used for all parameters are included in Table 2:

PSF	Range Sampled (inclusive)
PSA Score	*0 – 15*
Primary Gleason Score (pGleason)	*1 – 5*
Secondary Gleason Score (sGleason)	*1 – 5*
Total Gleason Score	*pGleason + sGleason*
Total Cores Sampled	*6 – 25*
Total Cores Positive	*0 – Total Cores Sampled*
ECOG Status	*1 – 5*
Stage	*1 – 4*
Age	*40 – 95*
Weight [kg]	*65 – 125*
Height [cm]	*150 – 200*
BMI [kg/m^2]	*Weight / (Height/100)^2*
Diabetes	*Boolean*
IBS	*Boolean*
COPD	*Boolean*
Implants	*Boolean*

Table 2: Sampling range for each PSF

In this model, each sampled PSF was made to be directly correlated with prostate motion. Most of the sampled PSFs were multiplied by a coefficient weight and that contribution

was added to the total motion. The coefficient weights were pre-determined and chosen such that the output motion was clinically meaningful. The Boolean PSFs contributed to motion depending on the result of the Boolean. Others contributed to motion if their sampled value was above some threshold tolerance value.

After determining the extent of the motion due to PSFs, noise was added to the motion. The noise was of the order of radiation therapist inter- and intra-fraction marker matching noise or approximately 0.85mm [37]. This was sampled from a Gaussian distribution with mean of zero and a standard deviation of 0.85mm. In reality, this value represents a minimum noise floor as the true noise is likely larger due to uncorrelated PSFs. The impact of noise on the performance of algorithms is examined in the results section.

A final step in creating this motion data was outlier data rejection and re-sampling. This was necessitated by BMI in particular. BMI was calculated from the height and weight data as described earlier, however by simply sampling from height and weight alone using a uniform distribution, unrealistic values of BMI are possible. To counteract this, BMI values exceeding 50 and no less than 10 were rejected and re-sampled.

The total extent of motion was restricted to between 0.01mm and 100mm. The lower limit was to ensure the evaluation of algorithm performance was not decreased by small motion precision errors. The upper limit is to ensure realistic motion; one would not expect the prostate to leave the patient's body during treatment for example.

Evaluation of Algorithms

Simulations that evaluated the algorithms ranged from 20, 50, and 200 simulated patients. Several common algorithms were tested using mostly default tuning parameters to identify candidate algorithms. The performance of each algorithm was evaluated using the distance between the predicted and actual motion (referred to as prediction error). The prediction error was plotted for each patient along an axis when evaluating the algorithm's performance. The mean, standard deviation, minimum, maximum, and mean absolute error (MAE) of the prediction error were also calculated. MAE in particular is a commonly used performance metric and used as a principle statistic.

As mentioned in the background section, scaling of data is critically important to the performance of some predictive algorithms. Therefore, all input data was scaled by removing the mean and dividing by the standard deviation for each feature:

$$x_i^{scaled} = \frac{\left(x_i^{unscaled} - \bar{x}_i\right)}{\sigma_{x_i}}$$

Where $\bar{x}_i$ is the mean value and σ_{x_i} is the standard deviation of a particular input feature, x_i.

As discussed in the background section, many of the algorithms are tuneable though the use of tuning parameters. This tuning is done based on the properties of the dataset in hand. All the algorithms were evaluated using default tuning parameters for the simulations.

Number of Patients Required

Multiple simulations were performed to estimate the number of patients required to predict patient intra-fraction motion based on their unique PSFs. Simulations were done for 20, 50, 100, 200, and 800 patients. Ten trials were run for each patient number to account for statistical fluctuations in algorithm performance caused by random sampling. The mean and standard deviation of the MAE for each trial was calculated for each patient number. This provides an estimate of the mean algorithm performance and the reproducibility of that performance for a given patient number. PSF sampling and motion generation was performed identically to that described in the prior section.

Effect of noise in the data

The largest uncertainty in this model was the amount of noise present in the simulated data. While noise estimates for the simulations were set to 0.85mm as described earlier, this represents a noise floor. The true noise in the data is likely greater due to errors in data collection and uncorrelated PSFs, for example. With this in mind, additional simulations were performed at different noise levels.

Simulations were completed as described in the previous section under different levels of Gaussian noise. The performance of algorithms was evaluated for noise values of 0.0mm, 0.4mm, 0.85mm, 1.5mm, 3.0mm, and 10.0mm, corresponding to the standard deviation of a Gaussian distribution, and three trials were performed for each noise value. PSF sampling and motion generation with new noise values was performed identically to that described in the motion generation section.

4.3.5 Applying Predictive Algorithms to Real Patient Data

The predictive algorithms were applied to the actual patient data using a methodology similar to that used in the benchmarking of the algorithms. Intra-fraction motion data from 21 patients along 3 directions were used for training and validation of the selected predictive algorithms. Each patient was associated with 16 PSFs, which were scaled and used as input to the machine learning algorithms.

Leave-One-Out cross validation was used to train and evaluate algorithm performance. The performance of the algorithms on the actual patient data was evaluated identically to their performance on the simulated patient data.

An important difference between real and simulated patient data is that real patient data often contains incomplete data. To address this, the missing fields were filled with the mean value of the particular PSF as shown in Table 3. Since the input data was scaled to remove the mean, this approach minimized the impact of the missing feature.

	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	MEAN
Raw PSF	4	Missing	6	5	Missing	5
Cleaned PSF	4	5*	6	5	5*	5

Table 3: Example of how incomplete data was filled for a PSF

4.4 Generating a Patient Specific Margin

4.4.1 Population Based Margin

The van Herk formula [6] was used to estimate a population-based PTV margin based on systematic and random errors in the radiation therapy process:

$$PTV\ Margin = 2.5\sum_{pop} + 0.7\sigma_{pop}$$

For detailed descriptions of these errors, refer to the radiation therapy background section or van Herk [6]. Numerical estimates for all these errors are included in Table 4. Note that inter-fraction motion errors were estimated to be 0.0mm because they were corrected for using image guidance.

Error Source	Vertical [mm]	Longitudinal [mm]	Lateral [mm]
Systematic Error	**1.4**	**1.6**	**1.0**
Target Delineation	0.7	1.3	0.7
Setup	0.5	0.5	0.5
Inter-fraction motion	0.0	0.0	0.0
Intra-fraction motion	1.0	0.9	0.5
Random Error	**1.9**	**1.8**	**1.2**
Setup	0.9	0.9	0.9
Inter-fraction motion	0.0	0.0	0.0
Intra-fraction motion	1.7	1.5	0.8

Table 4: Estimation of errors

4.4.2 Personalized Margin

A variation on the van Herk formula was used to generate a personalized margin. In this approach the population-based random error due to motion was replaced by a personalized random error due to motion. This is reasonable given that the population-based random motion error is simply an average all the individual random motion errors. More explicitly:

$$PTV\ Margin = 2.5\sum_{pop} + 0.7\sigma_{pers}$$

Where:

$$\Sigma_{\text{pop}}^2 = \Sigma_m^2 + \Sigma_s^2 + \Sigma_d^2$$

And:

$$\sigma_{pers}^2 = \sigma_{m,pers}^2 + \sigma_{s,pop}^2$$

$\Sigma_m, \Sigma_s, and\ \Sigma_d$ are the systematic errors due to motion, set-up, and delineation.

$\sigma_{s,pop}$ is a random set-up error which affects the entire population of patients.

$\sigma_{m,pers}$ is a personalized random error due based on predicted intra-fraction motion using PSFs.

5 Results

5.1 Data Analysis

5.1.1 Patient Specific Factor Statistics

PSFs were recorded for the 21 patients in this study. Table 5, Table 6, Table 7, and Table 8 provide statistics on the PSFs used for analysis.

Patient Specific Factor	Sample	Mean	Standard Deviation	Minimum	Maximum
PSA [ug/L]	*21*	*4.6*	*5.6*	*0.1*	*15.3*
Primary Gleason	*21*	*3.7*	*0.6*	*3*	*5*
Secondary Gleason	*21*	*4.0*	*0.7*	*3*	*5*
Total Gleason	*21*	*7.8*	*0.8*	*7*	*9*
Total Cores	*21*	*11.0*	*4.6*	*3*	*22*
Positive Cores	*21*	*7.1*	*3.5*	*1*	*12*
Age	*21*	*74.3*	*5.9*	*59*	*86*
Height [cm]	*21*	*171.6*	*10.3*	*139*	*185*
Weight [kg]	*21*	*80.9*	*11.0*	*58.4*	*105.1*
BMI [kg/m^2]	*21*	*27.7*	*4.8*	*19.5*	*42.1*

Table 5: PSF statistics

Number of Seeds	3	4	5
Number of Patients	*14*	*6*	*1*

Table 6: Number of implanted fiducial markers (seeds) patient distribution

ECOG Status	0	1	2	3	4	5	Missing
Number of Patients	*9*	*7*	*1*	*0*	*0*	*0*	*4*

Table 7: ECOG status patient distribution

Patient Specific Factor	Yes	No
Diabetes	*4*	*17*
IBS	*1*	*20*
COPD	*2*	*19*

Table 8: Boolean PSF statistics

5.1.2 Identification of Patient Specific Factors

Pearson and Spearman correlation tests revealed potential PSFs from the sample of 21 patients. Statistics investigated were a patient's mean intra-fraction motion and the standard deviation of intra-fraction motion along all three coordinate axis (Longitudinal, Vertical, Lateral). The purpose of identifying the PSFs in this section is to determine the most critical parameters in predicting a patient's intra-fraction motion. This process is done to reduce the dimensionality of the input data, which is an important step in improving predictive power and/or reducing the number of data points required for further work. Inter-fraction and intra-fraction rotation of fiducial markers was also investigated for PSF correlations. Fiducial marker rotation information was captured using 6D marker match.

Table 9 displays some of the Pearson and Spearman correlations between various PSFs and a patient's mean 3D intra-fraction motion. Displayed PSFs were chosen based on having absolute value of either Pearson or Spearman correlation greater than 0.3. Large differences between a Pearson and Spearman correlation for any given parameter can often be attributed to the influence of outlier points.

Pearson and Spearman Correlation between PSFs and mean 3D intra-fraction motion			
Direction	**PSF**	**Pearson**	**Spearman**
Longitudinal (Sup-Inf)	*ECOG Status*	*0.38*	*0.51*
	IBS	*-0.34*	*-0.33*
	Weight	*-0.21*	*-0.36*
Vertical (Ant – Post)	*IBS*	*0.40*	*0.37*
	ECOG Status	*-0.31*	*-0.29*
Lateral (Left – Right)	*BMI*	*-0.44*	*-0.16*
	Total Gleason	*-0.39*	*-0.44*
	Diabetes	*-0.34*	*-0.28*
	Secondary Gleason	*-0.28*	*-0.32*
	Height	*0.55*	*0.53*

Table 9: Pearson and Spearman correlations (>0.3) between mean 3D intra-fraction motion and PSFs

Using the same criteria, Table 10 shows correlations between PSFs and a patient's standard deviation of intra-fraction motion.

Pearson and Spearman Correlation between PSFs and std of 3D intra-fraction motion			
Direction	**PSF**	**Pearson**	**Spearman**
Longitudinal (Sup-Inf)	*PSA Score*	*-0.41*	*-0.50*
	Primary Gleason	*-0.38*	*-0.36*
	COPD	*0.24*	*0.27*
	Secondary Gleason	*0.28*	*0.31*
	IBS	*0.43*	*0.33*
Vertical (Ant – Post)	*PSA Score*	*-0.38*	*-0.34*
	Height	*0.24*	*0.30*
	Weight	*0.30*	*0.21*
	IBS	*0.48*	*0.37*
Lateral (Left – Right)	*Weight*	*-0.36*	*-0.17*
	PSA Score	*0.38*	*0.21*

Table 10: Pearson and Spearman correlations (>0.3) between the standard deviation of 3D intra-fraction motion and PSFs

While the BCCA-VIC performs all couch shifts for prostate cancer treatment using only 3D shifts, the marker matching software allows for a 6D shift to be calculated. The 6D shift allows for prostate rotation to be accounted for in addition to the standard couch shift. Table 11 and Table 12 show the Pearson and Spearman correlations between the mean and standard deviation of intra-fraction rotations and the PSFs. The intra-fraction rotations were calculated as the difference between the post-image rotation and the pre-image rotation.

Pearson and Spearman Correlation between PSFs and mean 6D intra-fraction rotation			
Direction	**PSF**	**Pearson**	**Spearman**
Pitch	*Weight*	*-0.09*	*-0.32*
	Positive Cores	*-0.29*	*-0.32*
	IBS	*-0.19*	*-0.30*
	Primary Gleason	*0.30*	*0.38*
	Total Cores	*0.33*	*0.27*
	Total Gleason	*0.13*	*0.30*
Roll	*Diabetes*	*-0.46*	*-0.29*
	Primary Gleason	*-0.31*	*-0.27*
	IBS	*0.32*	*0.34*
	Total Gleason	*-0.30*	*-0.32*
	PSA Score	*0.25*	*0.35*
Yaw	*Age*	*-0.40*	*-0.19*
	Secondary Gleason	*-0.31*	*-0.15*
	PSA Score	*0.30*	*0.42*

Table 11: Pearson and Spearman correlation between mean 6D intra-fraction rotation and PSFs

Pearson and Spearman Correlation between PSFs and the std of 6D intra-fraction rotation			
Direction	**PSF**	**Pearson**	**Spearman**
Pitch	*PSA Score*	*-0.30*	*-0.38*
	Number of Fiducials	*-0.12*	*-0.41*
	COPD	*-0.25*	*-0.41*
Roll	*ECOG Status*	*-0.48*	*-0.50*
	Number of Fiducials	*-0.36*	*-0.48*
	Weight	*-0.34*	*-0.45*
	BMI	*-0.33*	*-0.28*
	PSA Score	*-0.31*	*-0.35*
	Secondary Gleason	*0.30*	*0.27*
	Diabetes	*0.50*	*0.37*
	Total Cores	*-0.20*	*-0.55*
Yaw	*Weight*	*-0.38*	*-0.28*
	Positive Cores	*-0.28*	*-0.47*
	Diabetes	*-0.25*	*-0.30*
	PSA Score	*-0.26*	*-0.30*
	ECOG Status	*0.27*	*0.38*
	BMI	*-0.27*	*-0.30*

Table 12: Pearson and Spearman correlation between the standard deviation of 6D intra-fraction rotation and PSFs

Finally, the correlations between the mean and standard deviation of inter-fraction rotations and PSFs are presented in Table 13 and Table 14. The inter-fraction rotations are not corrected for prior to treatment and may therefore be important.

Pearson and Spearman Correlation between PSFs and mean 6D inter-fraction rotation			
Direction	**PSF**	**Pearson**	**Spearman**
Pitch	*Primary Gleason*	*-0.43*	*-0.39*
	Total Gleason	*-0.42*	*-0.47*
	Age	*0.38*	*0.17*
	COPD	*0.24*	*0.31*
Roll	*BMI*	*-0.55*	*-0.40*
	Positive Cores	*-0.46*	*-0.52*
	PSA Score	*-0.43*	*-0.32*
	Number of Fiducials	*-0.32*	*-0.25*
	Weight	*-0.31*	*-0.30*
	Height	*0.32*	*0.23*
Yaw	*Number of Fiducials*	*-0.54*	*-0.58*
	Total Cores	*-0.17*	*-0.42*
	Height	*-0.14*	*-0.31*

Table 13: Pearson and Spearman correlation between mean 6D inter-fraction rotation and PSFs

Pearson and Spearman Correlation between PSFs and the std of 6D inter-fraction rotation			
Direction	**PSF**	**Pearson**	**Spearman**
Pitch	*Positive Cores*	*-0.39*	*-0.28*
	COPD	*-0.36*	*-0.47*
	ECOG Status	*0.32*	*0.48*
	Diabetes	*-0.29*	*-0.34*
	Primary Gleason	*0.25*	*0.32*
	Height	*0.21*	*0.41*
Roll	*Number of Fiducials*	*-0.38*	*-0.40*
	Positive Cores	*-0.30*	*-0.33*
	Diabetes	*0.34*	*0.24*
	Secondary Gleason	*0.37*	*0.39*
	Total Gleason	*0.39*	*0.33*
	PSA Scores	*-0.08*	*-0.43*
	Total Cores	*-0.25*	*-0.40*
	COPD	*-0.27*	*-0.38*
	ECOG Status	*-0.19*	*-0.32*
Yaw	*Positive Cores*	*-0.40*	*-0.46*
	Total Cores	*-0.36*	*-0.17*
	BMI	*-0.20*	*-0.43*
	Diabetes	*-0.18*	*-0.40*
	Secondary Gleason	*-0.07*	*-0.38*
	ECOG Status	*0.27*	*0.50*
	Primary Gleason	*0.20*	*0.52*
	PSA Score	*0.39*	*-0.01*

Table 14: Pearson and Spearman correlation between the standard deviation of 6D inter-fraction rotation and PSFs

5.1.3 Time Dependent Data

The time dependent data was defined as data that varies on a fraction-to-fraction basis for each patient. This includes motion and rotation data collected each fraction, along with data about the presence of gas, and fraction timestamp information. The only correlation of significance here was between longitudinal intra-fraction motion and vertical intra-fraction motion (Table 15). These parameters are strongly correlated with each other, and may be due to pelvic tilt.

Correlation between Time Dependent Parameters (3D)			
Time Dependent Parameter 1	**Time Dependent Parameter 2**	**Pearson**	**Spearm**
Vertical intra-fraction motion	Longitudinal intra-fraction motion	*-0.74*	*-0.70*

Table 15: Correlation between vertical and longitudinal intra-fraction motion

Also notable here is the lack of correlation between motion and time stamp information, as previous studies have suggested longer treatment times may be associated with increased intra-fraction motion [33,38,39].

The 6D fiducial marker rotations showed no meaningful correlations between time dependent parameters.

5.1.4 Rectal Distention

The mean intra-fraction motion is affected by changes in gas during treatment. A change in gas means either gas is not present prior to the fraction, but is present after the fraction (N-Y), or the vice-versa (Y-N). This is explained in detail in the methodology section.

Fractions where a change in gas occurred represent a very small proportion of the total number of fractions (~ 4%).

Table 16 and Table 17 show the effect intra-fractional gas change on the mean and standard deviation of intra-fraction motion, respectively.

Mean Intra-fraction Motion				
	N-N	**N-Y**	**Y-N**	**Y-Y**
Fractions	*502*	*26*	*1*	*122*
Lng [mm]	*0.27*	*-0.91*	*-4.9*	*0.12*
Vrt [mm]	*-0.77*	*0.21*	*4.4*	*-0.63*
Lat [mm]	*0.08*	*0.13*	*0.2*	*0.05*

Table 16: Effect of gas on mean intra-fraction 3D motion

Standard Deviation of Intra-fraction Motion				
	N-N	**N-Y**	**Y-N**	**Y-Y**
Fractions	*502*	*26*	*1*	*122*
Lng [mm]	*1.66*	*1.78*	*N/A*	*1.79*
Vrt [mm]	*1.84*	*1.72*	*N/A*	*2.17*
Lat [mm]	*1.02*	*0.68*	*N/A*	*0.78*

Table 17: Effect of gas on the standard deviation of intra-fraction 3D motion

A change in gas appears to have a significant (>0.5mm) effect on mean intra-fraction motion along both the longitudinal and vertical axes. There is little effect along the lateral axis. The standard deviation of intra-fraction motion is not affected by gas to the same degree as mean intra-fraction motion.

Table 18 and Table 19 show the effect of pre-treatment gas on the mean and standard deviation of inter-fraction rotation, respectively. This rotation is not corrected for prior to treatment and consequentially may be important.

Mean Inter-fraction Rotation		
	Pre N	**Pre Y**
Fractions	*542*	*127*
Pitch [deg]	*-6.9*	*-4.2*
Roll [deg]	*0.3*	*1.5*
Yaw [deg]	*0.1*	*0.2*

Table 18: Effect of pre-treatment gas on mean inter-fraction rotation

Standard Deviation of Inter-fraction Rotation		
	Pre N	**Pre Y**
Fractions	*542*	*127*
Pitch [deg]	*8.9*	*10.2*
Roll [deg]	*3.0*	*3.4*
Yaw [deg]	*4.6*	*4.8*

Table 19: Effect of pre-treatment gas on the standard deviation of inter-fraction rotation

There are some rotation differences between fractions with/without pre-treatment gas. Pre-treatment gas appears to increase the mean roll deviation from the treatment plan, but decrease the mean pitch deviation. The standard deviations of rotation is larger with pre-treatment gas along all rotational axes, but most notably along the pitch axis.

Table 20 and Table 21 show the effect of a change in gas during treatment on the mean and standard deviation of intra-fraction rotation, respectively. Intra-fraction rotations were

calculated as the difference between the post-treatment rotation correction and pre-treatment rotation correction, as the rotations are not corrected for prior to treatment.

Mean Intra-fraction Rotation				
	N-N	**N-Y**	**Y-N**	**Y-Y**
Fractions	*502*	*26*	*1*	*122*
Pitch [deg]	*1.1*	*-2.9*	*1.4*	*3.7*
Roll [deg]	*0.1*	*0.3*	*-5.8*	*-0.1*
Yaw [deg]	*0.2*	*-0.2*	*-0.4*	*0.0*

Table 20: Effect of a change in gas during treatment on mean intra-fraction rotation

Standard Deviation of Intra-fraction Rotation				
	N-N	**N-Y**	**Y-N**	**Y-Y**
Fractions	*502*	*26*	*1*	*122*
Pitch [deg]	*5.0*	*4.9*	*N/A*	*6.8*
Roll [deg]	*2.8*	*2.5*	*N/A*	*2.0*
Yaw [deg]	*1.7*	*1.4*	*N/A*	*2.3*

Table 21: Effect of a change in gas during treatment on the standard deviation of intra-fraction rotation

There is a slight change in the mean intra-fraction pitch during treatment due to a change in gas. This is expected since the presence of gas is generally in the rectum, below the prostate. Mean intra-fraction roll and yaw are not affected to the same degree. Overall, there is little change in the standard deviation of intra-fraction rotation due to intra-fractional gas change.

5.1.5 Population-Based Statistics

Population-based motion statistics were compiled in order to gauge an average patient's prostate intra-fraction motion. Table 22 shows the population average intra-fraction prostate motion. Motion was most prevalent along the vertical and longitudinal axes, as expected. In particular, the mean vertical motion showed a significant deviation along the vertical axis from planned location. Motion was considered significant if it is greater than 0.5mm, which is the limit on precision of the couch motion.

Population-based statistics for 3D intra-fraction motion			
	Vertical	**Longitudinal**	**Lateral**
Mean [mm]	*-0.70*	*0.19*	*0.08*
Standard Deviation [mm]	*1.92*	*1.71*	*0.96*

Table 22: Mean and standard deviation of intra-fraction motion for all 678 fractions

To investigate the importance of motion asymmetry along any axis, the directionality of the motion (ex: positive vs negative longitudinal motion) was investigated (Table 23). There was a greater extent of mean intra-fraction motion posteriorly compared to anteriorly. A similar posterior drift of the prostate was found by others, who used real-time imaging, as well [39,40].

Directionality of 3D intra-fraction motion						
	Vrt (A)	**Vrt (P)**	**Lng (I)**	**Lng (S)**	**Lat (R)**	**Lat (L)**
Mean [mm]	*1.25*	*-1.65*	*1.12*	*-1.24*	*0.65*	*-0.73*
Standard Deviation [mm]	*1.37*	*1.34*	*1.15*	*1.42*	*0.67*	*0.69*

Table 23: Directionality of intra-fraction motion. Mean and standard deviation of intra-fraction motion along the positive and negative directions for each axis

Population-based statistics were also compiled for rotations using 6D data. Table 24 and Table 25 show the mean and standard deviation of inter-fraction and intra-fraction rotations, respectively. The most significant rotation was the pitch, as expected. This corresponds to rotation about the lateral axis, the natural rotation axis of the pelvis. Pitch rotation may be influenced by a number of factors including variability in bladder fullness, rectal distension, and positioning on the couch. There was generally greater rotational error caused by inter-fraction rotation than intra-fraction rotation.

Population-based statistics for 6D inter-fraction rotation			
	Pitch	**Roll**	**Yaw**
Mean [deg]	*-6.4*	*0.5*	*0.1*
Standard Deviation [deg]	*9.2*	*3.1*	*4.6*

Table 24: Mean and standard deviation of inter-fraction rotation for all 678 fractions

Population-based statistics for 6D intra-fraction rotation			
	Pitch	**Roll**	**Yaw**
Mean [deg]	*1.4*	*0.1*	*0.2*
Standard Deviation [deg]	*5.5*	*2.7*	*1.8*

Table 25: Mean and standard deviation of intra-fraction rotation for all 678 fractions

The directionality of inter-fraction and intra-fraction rotation also reveals a large range of pitch rotation (Table 26 and Table 27). The extent of inter-fraction pitch along the negative direction is particularly large with a mean value of -10.0°. Interestingly, this is contrasts intra-fraction pitch rotation, which has a larger extent of pitch rotation in the positive direction (4.1° vs -3.3°). Overall, there is again a greater extent of rotation error inter-fractionally than intra-fractionally.

Directionality of 6D inter-fraction rotation						
	Pitch (+)	**Pitch (-)**	**Roll (+)**	**Roll (-)**	**Rot (+)**	**Rot (-)**
Mean [deg]	*4.8*	*-10.0*	*2.7*	*-2.1*	*3.0*	*-2.5*
Standard Deviation [deg]	*4.5*	*7.3*	*2.3*	*1.4*	*5.1*	*1.8*

Table 26: Directionality of patient rotation. Mean and standard deviation of inter-fraction rotation along the positive and negative directions for each axis

Directionality of 6D intra-fraction rotation						
	Pitch (+)	**Pitch (-)**	**Roll (+)**	**Roll (-)**	**Rot (+)**	**Rot (-)**
Mean [deg]	*4.1*	*-3.3*	*1.5*	*-1.5*	*1.2*	*-1.2*
Standard Deviation [deg]	*4.2*	*4.2*	*2.5*	*1.8*	*1.2*	*1.6*

Table 27: Directionality of patient motion. Mean and standard deviation of intra-fraction rotation along the positive and negative directions for each axis

5.2 Predictive Algorithm Benchmarking

5.2.1 Number of Patients Required

Using the simulated patient motion data, the feasibility of the various machine learning algorithms were investigated for potential use in predicting patient intra-fraction motion. Algorithms were evaluated using leave-one-out cross validation with default tuning parameters, as described in the methodology. Figure 7 shows the prediction mean absolute error (MAE) as a function of the number of simulated patients. Plotted MAE represents the mean value of 10 MAEs calculated on 10 independent trials. Error bars were omitted for clarity, but included in a duplicate plot within the appendix.

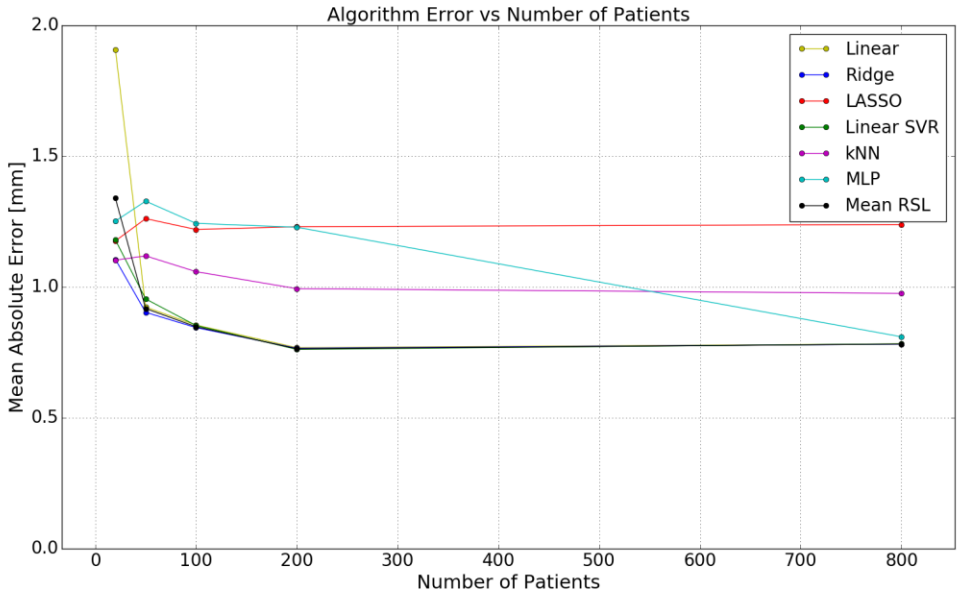

Figure 7: Projected performance of algorithms based on simulated data

The models generally improved their performance with increased data. In particular, there was a large improvement in the performance of most of the models within the first 50-100 patients. The algorithm's performance also seemed to stop improving around the noise floor of 0.85mm in the simulated data.

Models such as the linear model (which are more sensitive to noise), showed an even larger improvement in the earlier stages of data collection. In contrast to the linear model, regularized models, such as the Ridge and LASSO, were far more robust at lower numbers of sample data points. This robustness with small amounts of data generally comes at the expense of accuracy with large amounts of data. This was seen here for the LASSO, but not the Ridge model.

The MLP was a more complicated model, as it had a large number of parameters to be trained. Consequently, it required a larger sample of data before it improved to the same degree as the other models. It was the only model to see a significant improvement beyond 200 patients.

The kNN is a simple model that makes predictions by association. It performed well with few patients available, likely due to the similarity in motion between all patients, but it wasn't able to improve its performance much with larger amounts of data.

The linear SVR (SVR) attempted to find a linear solution such that none of the training data predictions differ from the true value by more than some user-specified threshold value. This value was set to 0.5mm, which reflects the resolution of margin definition in the treatment planning system, as explained in the methodology section. The SVR performed very well throughout and outperformed all others examined, except for the Ridge.

The mean RSL was an ensemble algorithm with equal influence from the Linear, Ridge, and SVR models. It performed well, but not noticeably better than the Ridge or SVR independently.

5.2.2 Evaluation of Algorithms

Separate simulations using a single trial were done to provide a more detailed breakdown of algorithm performance. These simulations were performed for all algorithms, however only the Linear model, the Ridge model, and the MLP model are included here. Plots such

as those in Figure 8 to Figure 10 are useful to gauge how an algorithm is performing beyond a single metric and may be useful in future quality assurance protocols designed for statistical learning algorithms.

The Linear algorithm showed the most improvement within the first 50-100 patients. A detailed breakdown of its performance is included in Figure 8, where it again shows a large degree of error with only 20 patients available, but improves its performance in simulations with larger numbers of patients.

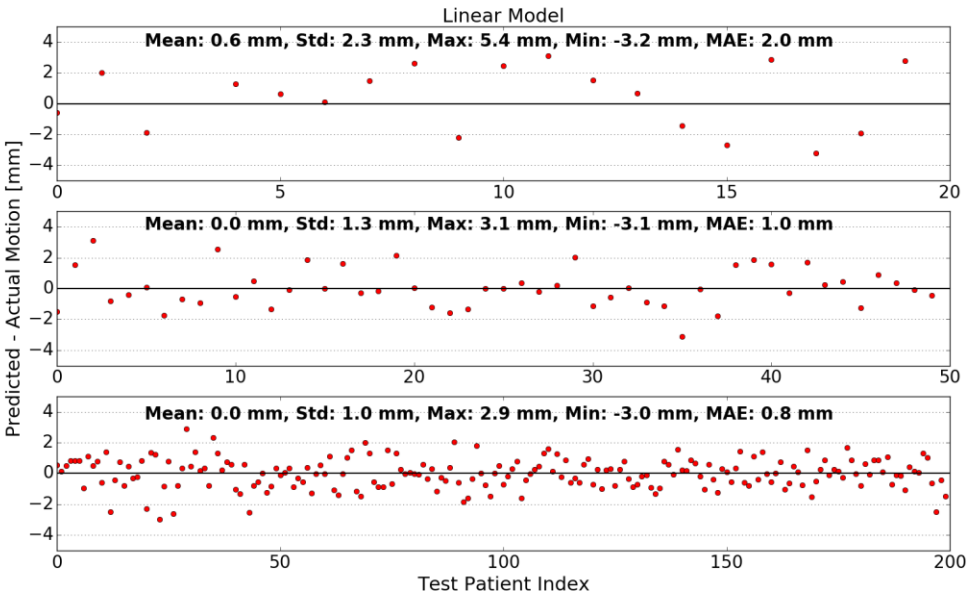

Figure 8: Performance of the Linear model using simulated data from 20, 50, and 200 patients.

The Ridge algorithm demonstrated the best performance overall in the simulations. Figure 9 shows a detailed breakdown of its performance for a single trial. Compared to the Linear model, it shows improved performance in the lower number of patients regime, but performs similar to the Linear model at larger number of patients.

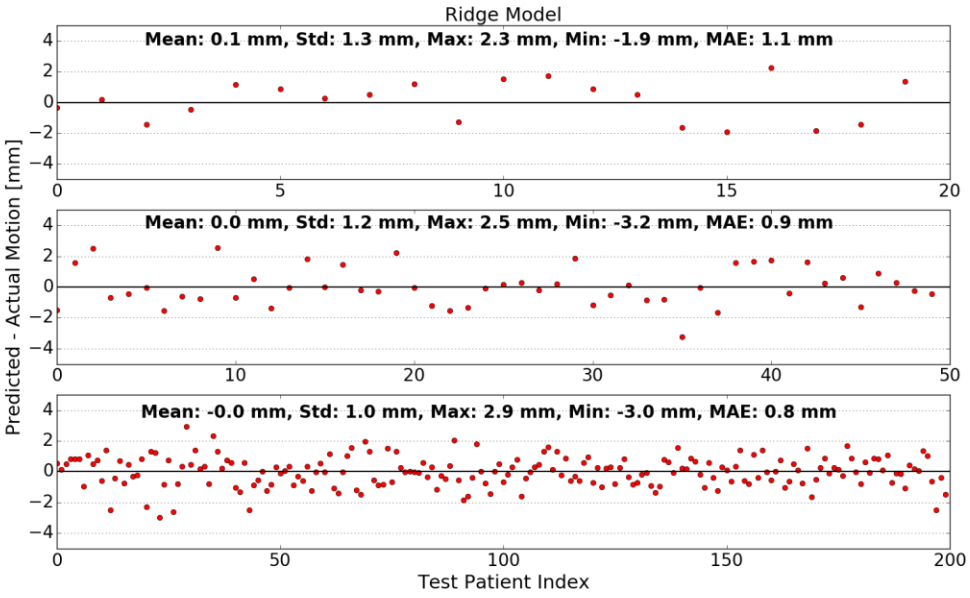

Figure 9: Performance of the Ridge model using simulated data from 20, 50, and 200 patients.

Figure 10 shows the performance of the MLP algorithm for one trial. The MLP previously showed the greatest improvement with larger amounts of data. Here the MLP model performed relatively well early on, but didn't improve its performance appreciably with 200 patients.

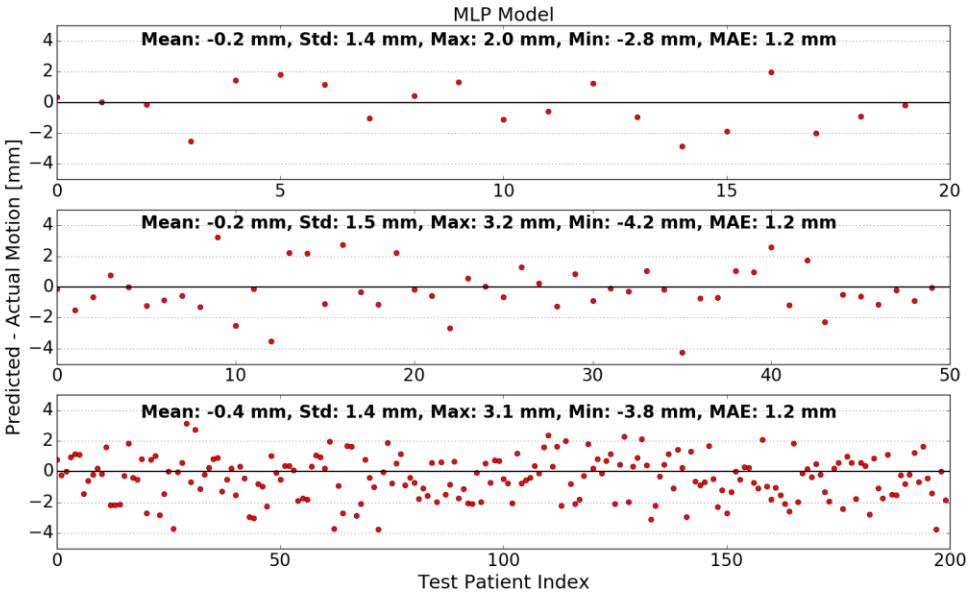

Figure 10: Performance of the MLP model using simulated data from 20, 50, and 200 patients.

5.2.3 Influence of Noise

While the noise was estimated to be 0.85mm based on the radiation therapist inter-observer and intra-observer marker matching variability, the noise may differ from that amount in practice. It is therefore useful to see how these algorithms perform under different amounts of noise. At the very least, if the algorithms can improve their performance in the presence of a lot of noise, and if PSFs are at least partially relevant to intra-fraction motion, then the algorithms can be expected to improve their performance with real patient data. Some of the algorithms which seemed more useful were also investigated for the effects of different levels of noise in the data. Error bars were again omitted for clarity, but identical plots with error bars are available in the appendix.

Figure 11 shows the performance of the linear model at different levels of Gaussian patient motion noise. Overall, the linear model performed poorly in the presence of greater noise, likely the result of failing to distinguish between input data noise and underlying signal in the data.

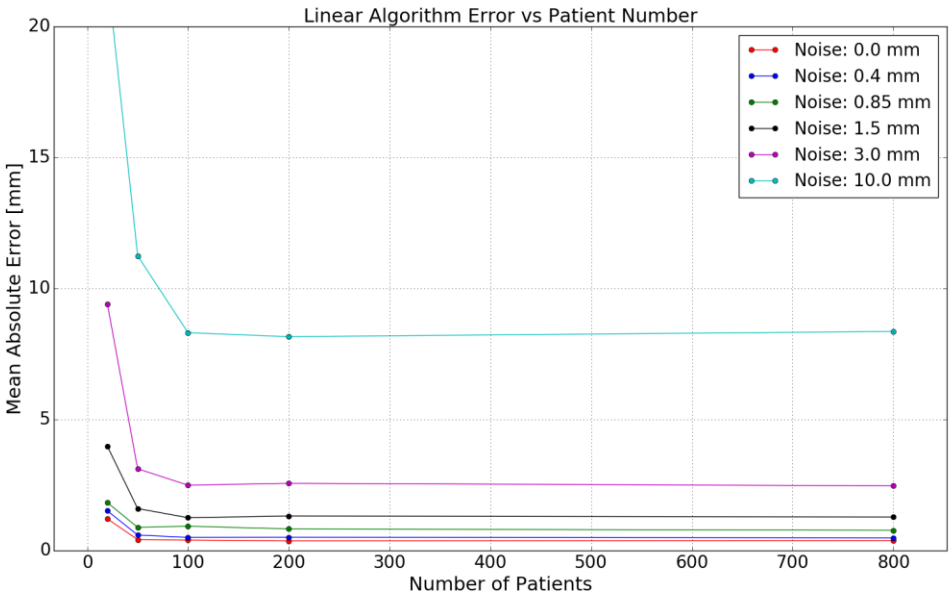

Figure 11: Effect of noise on linear model

Figure 12 shows the effect of noise on the Ridge model. It performed better than the linear model with increased noise, especially with fewer patients.

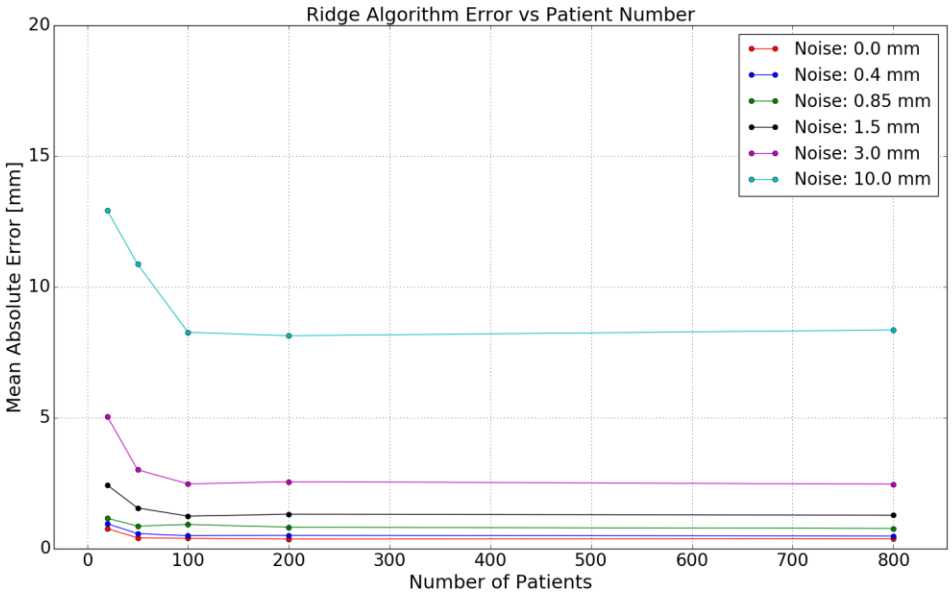

Figure 12: Effect of noise on ridge model

Finally, Figure 13 shows the effect of noise on the MLP model. It showed a more steady improvement, but was still hampered by the increased noise.

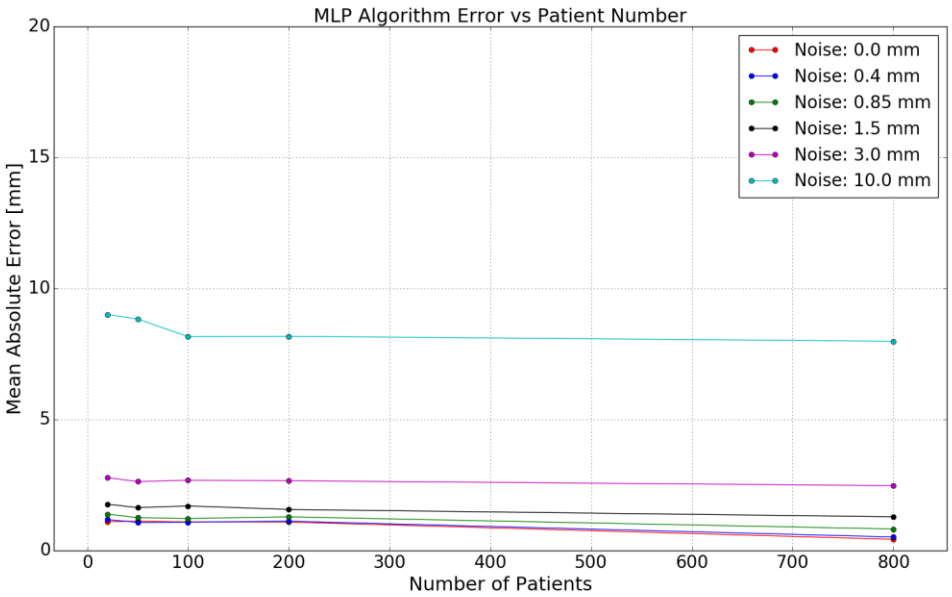

Figure 13: Effect of noise on the MLP model

5.3 Predictive Algorithms Applied to Real Patient Data

5.3.1 Intra-fraction Motion

An identical methodology to the previous section was used on the sample of 21 patients. The LASSO model was selected based on having the lowest MAE along all of the three axis; however, the choice of algorithm could change with higher quantities of data. The algorithm was configured to predict both the mean and the standard deviation of a patient's intra-fraction motion.

Figure 14 shows that the LASSO model was able to predict the mean intra-fraction motion of the prostate to within 0.8mm MAE along all three directional axes. The vertical axis had the largest error whereas the lateral axis had the smallest error. This is not surprising given that the vertical axis had the largest standard deviation in mean intra-fraction motion whereas the lateral axis had the smallest (Table 22).

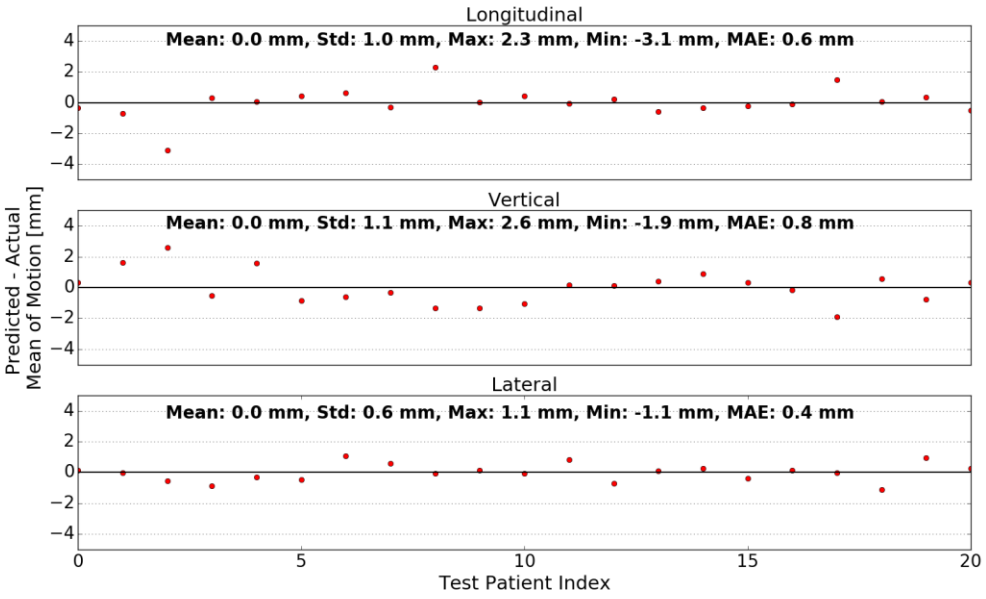

Figure 14: LASSO algorithm prediction of a patient's mean intra-fraction motion

Figure 15 shows that the LASSO was able to predict the standard deviation of the intra-fraction motion of the prostate to within 0.4mm MAE along all three directional axes. The standard deviation of intra-fraction motion was predicted to a higher accuracy than the mean intra-fraction motion.

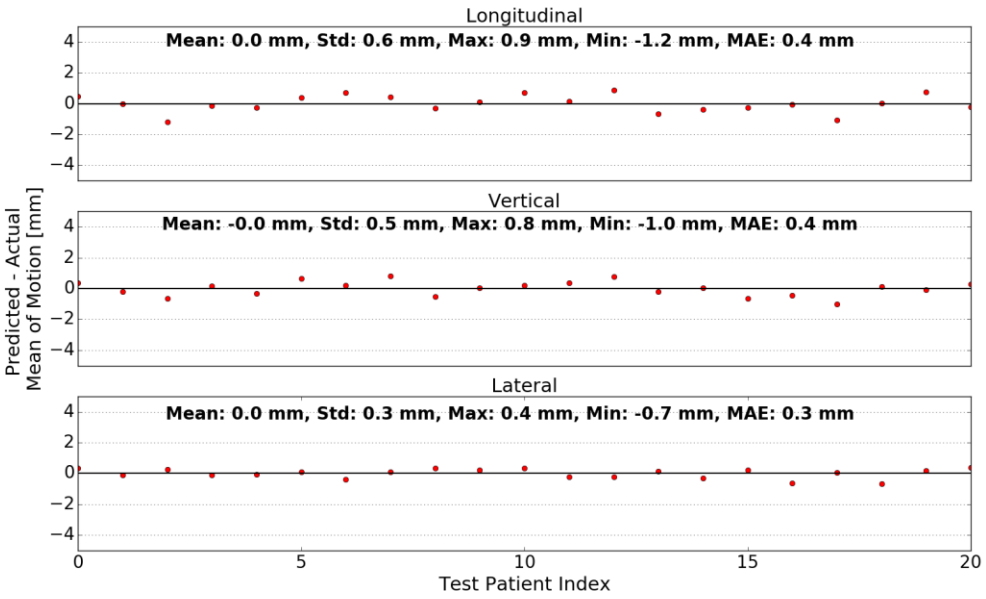

Figure 15: LASSO algorithm prediction of the standard deviation of a patient's intra-fraction motion

5.3.2 Algorithm Coefficient Weighting

Beyond the principle predictions produced by the algorithms, the coefficient weights assigned by the algorithm to each PSF were extracted. These were useful in order to gain insight into which PSFs are valued by the algorithm and will be helpful for feature selection when more data becomes available. A larger PSF coefficient means that an algorithm's prediction was influenced more strongly by that particular PSF. A smaller PSF coefficient means that an algorithm's prediction was influenced less strongly by that particular PSF.

Table 28 and Table 29 provide the PSF coefficients used for the LASSO model and Ridge model's mean intra-fraction motions and standard deviations of intra-fraction motion predictions, respectively. The LASSO had the best performance of all the models and its PSF coefficients were all exactly zero. This may be an early indicator that the PSFs may not be very helpful in predicting intra-fraction motion; however more data would be required to validate this. To provide some contrast, the magnitude of the PSF coefficients from the Ridge model are also included for the mean (Table 28) and standard deviation (Table 29) predictions as well. The Ridge model's PSF coefficients are clearly greater than zero in magnitude and, for n=21 patients, the Ridge model also performed more poorly than the LASSO model on the real patient data for both mean (Table 30) and standard deviation (Table 31).

Mean Intra-fraction Motion Coefficients						
	Longitudinal		**Vertical**		**Lateral**	
PSF Coefficient	**LASSO**	**Ridge**	**LASSO**	**Ridge**	**LASSO**	**Ridge**
PSA Score	*0*	*-0.0996*	*0*	*0.3529*	*0*	*0.0142*
Primary Gleason	*0*	*0.0572*	*0*	*-0.0865*	*0*	*-0.0072*
Secondary Gleason	*0*	*0.0253*	*0*	*0.1094*	*0*	*-0.0715*
Total Gleason	*0*	*0.0590*	*0*	*0.0297*	*0*	*-0.0625*
Positive Cores	*0*	*-0.2873*	*0*	*0.0328*	*0*	*0.0955*
Total Cores	*0*	*0.1000*	*0*	*-0.0716*	*0*	*0.0439*
Age	*0*	*0.1971*	*0*	*-0.0898*	*0*	*0.0940*
Weight	*0*	*0.0424*	*0*	*0.0103*	*0*	*0.0267*
Diabetes	*0*	*0.0426*	*0*	*-0.2514*	*0*	*-0.1903*
IBS	*0*	*-0.1817*	*0*	*0.2848*	*0*	*-0.0196*
COPD	*0*	*-0.0536*	*0*	*0.1774*	*0*	*0.1243*
Implants	*0*	*0*	*0*	*0*	*0*	*0*
ECOG Status	*0*	*0.3922*	*0*	*-0.4099*	*0*	*-0.0550*
BMI	*0*	*-0.0474*	*0*	*0.0874*	*0*	*-0.0803*
Height	*0*	*-0.1054*	*0*	*0.0919*	*0*	*0.2024*
IGRT	*0*	*-0.1072*	*0*	*0.0284*	*0*	*-0.0714*

Table 28: PSF coefficients used in mean intra-fraction motion predictions for LASSO and Ridge models.

Standard Deviation of Intra-fraction Motion Coefficients						
	Longitudinal		Vertical		Lateral	
PSF Coefficient	**LASSO**	**Ridge**	**LASSO**	**Ridge**	**LASSO**	**Ridge**
PSA Score	*0*	*-0.0995*	*0*	*-0.0414*	*0*	*0.1540*
Primary Gleason	*0*	*-0.1672*	*0*	*-0.0392*	*0*	*0.0145*
Secondary Gleason	*0*	*0.2110*	*0*	*0.1547*	*0*	*-0.0036*
Total Gleason	*0*	*0.0571*	*0*	*0.0981*	*0*	*0.0069*
Positive Cores	*0*	*-0.1959*	*0*	*-0.2100*	*0*	*0.0210*
Total Cores	*0*	*0.0710*	*0*	*-0.0014*	*0*	*-0.0672*
Age	*0*	*0.1139*	*0*	*0.0716*	*0*	*0.1015*
Weight	*0*	*0.0811*	*0*	*0.1094*	*0*	*-0.0878*
Diabetes	*0*	*-0.1061*	*0*	*-0.0298*	*0*	*-0.0020*
IBS	*0*	*0.1871*	*0*	*0.1812*	*0*	*0.0277*
COPD	*0*	*0.1373*	*0*	*0.0692*	*0*	*-0.0736*
Implants	*0*	*0*	*0*	*0*	*0*	*0*
ECOG Status	*0*	*0.0525*	*0*	*-0.0267*	*0*	*-0.0041*
BMI	*0*	*-0.0239*	*0*	*0.0367*	*0*	*-0.0158*
Height	*0*	*0.0690*	*0*	*0.1449*	*0*	*-0.0578*
IGRT	*0*	*0.1016*	*0*	*0.0631*	*0*	*-0.0466*

Table 29: PSF coefficients used in standard deviation of intra-fraction motion predictions for LASSO and Ridge models.

Mean Intra-fraction motion			
	Longitudinal	**Vertical**	**Lateral**
LASSO MAE [mm]	*0.6*	*0.8*	*0.4*
Ridge MAE [mm]	*1.5*	*1.6*	*0.6*

Table 30: Prediction error (MAE) comparison between LASSO and Ridge models for mean intra-fraction motion

Standard Deviation of Intra-fraction motion			
	Longitudinal	**Vertical**	**Lateral**
LASSO MAE [mm]	*0.4*	*0.4*	*0.3*
Ridge MAE [mm]	*0.7*	*0.7*	*0.4*

Table 31: Prediction error (MAE) comparison between LASSO and Ridge models for standard deviation of intra-fraction motion

A final note on the PSF coefficients is that the degree to which the LASSO model penalizes PSF coefficients can be changed by a tuneable parameter, α. This parameter tuning was discussed in detail in the statistics background section. Figure 16 provides an example of how parameter tuning can be performed with the LASSO model. This shows that, at least for n=21 patients, the smaller the PSF coefficients, the better the performance. A larger α forces the magnitude of the PSF coefficients to be smaller.

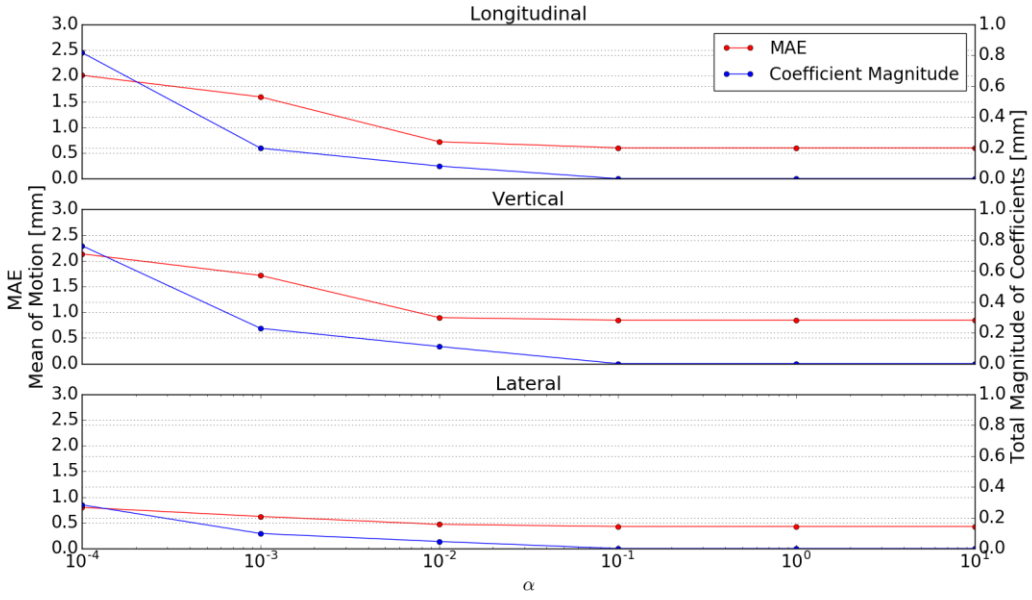

Figure 16: LASSO MAE and magnitude of PSF coefficients plotted against choice of alpha tuneable parameter

5.4 Generating Margins

5.4.1 Population-Based Margin

Using population-based motion statistics as well as other center-specific radiation therapy set-up and delivery uncertainties a population-based margin was generated using van Herk approach [6] outlined in the methodology. Table 32 provides estimates for these errors along with the associated PTV margin.

Error Source	Vertical [mm]	Longitudinal [mm]	Lateral [mm]
Systematic Error	**1.3**	**1.7**	**1.0**
Target Delineation	0.7	1.3	0.7
Setup	0.5	0.5	0.5
Inter-fraction motion	0.0	0.0	0.0
Intra-fraction motion	1.0	0.9	0.5
Random Error	**1.9**	**1.8**	**1.2**
Setup	0.9	0.9	0.9
Inter-fraction motion	0.0	0.0	0.0
Intra-fraction motion	1.7	1.5	0.8
Population-based PTV Margin			
Systematic PTV Component	3.3	4.2	2.5
Random PTV Component	1.4	1.2	0.8
PTV Margin	4.7	5.4	3.3

Table 32: Population-based PTV margin and errors

Currently at VIC, a 7mm uniform margin is used with the potential for reduction down to 5mm in the direction of the rectum if required when using kV-kV seed matching. Using this approach, it appears as though that PTV margin can be reduced from 7mm.

5.4.2 Personalized Margin

A variation on the van Herk formula was used to generate a personalized margin based on the predictions of intra-fraction motion combined with center-specific system and random uncertainties in treatment set-up and delivery.

$$PTV\ Margin = 2.5\sum_{pop} + 0.7\sigma_{pers}$$

Where Σ_{pop} is the total population-based systematic error and σ_{pers} is the total personalized random error. Σ_{pop} is composed of error contributions caused by systematic error in delineation, set-up, and motion. σ_{pers} is composed of error contributions caused by random population-based error in set-up and random personalized error in motion. While the mean personalized intra-fraction motion, M_{pers}, is also predicted using the algorithms, this was not considered in generating a personalized PTV margin since M_{pers} corresponds to a couch shift, rather than an uncertainty. However, this could be potentially useful for identifying patients that require more resources (time, imaging, tracking) because of their pre-disposition to large systematic motion.

Table 33 provides estimates for these errors along with the associated PTV margin for one patient. For n=21 patients, the personalized margin is identical to the population-based margin because the best algorithm ignored all input due to PSFs. However it has the opportunity to be personalized if PSFs prove to be more impactful on motion with more data.

Error Source	Vertical [mm]	Longitudinal [mm]	Lateral [mm]
Systematic Error	**1.3**	**1.7**	**1.0**
Target Delineation	0.7	1.3	0.7
Setup	0.5	0.5	0.5
Inter-fraction motion	0.0	0.0	0.0
Intra-fraction motion	1.0	0.9	0.5
Random Error	**1.9**	**1.8**	**1.2**
Setup	0.9	0.9	0.9
Inter-fraction motion	0.0	0.0	0.0
Intra-fraction motion	1.7	1.5	0.8
Personalized PTV Margin			
Systematic PTV Component	3.3	4.2	2.5
Random PTV Component	1.4	1.2	0.8
PTV Margin	4.7	5.4	3.3
Treatment Planning PTV	5	6	4

Table 33: Personalized PTV margin for one patient

Two plans were generated for this patient. One used the prior PTV margin and one used the newly calculated PTV margin. The prescription dose for both plans was 78 Gy delivered over 39 fractions and planned such that 100% of the prescription dose covers 98% of the PTV. The new treatment planning PTV margin (Table 33) was calculated by rounding-up the personalized PTV margin to the nearest integer. This was done because the treatment planning system only allows integer value margins. The two plans used identical optimization.

Figure 17 shows the two PTV margins and the treatment planning dose distributions. The new PTV is the plan on the left. The previous PTV plan is the one on the right. The new plan was acceptable for minimum dose (96.6%), maximum dose (103.3%), and uniformity.

The new plan achieves the same prescription dose to 98% of the volume as the original plan.

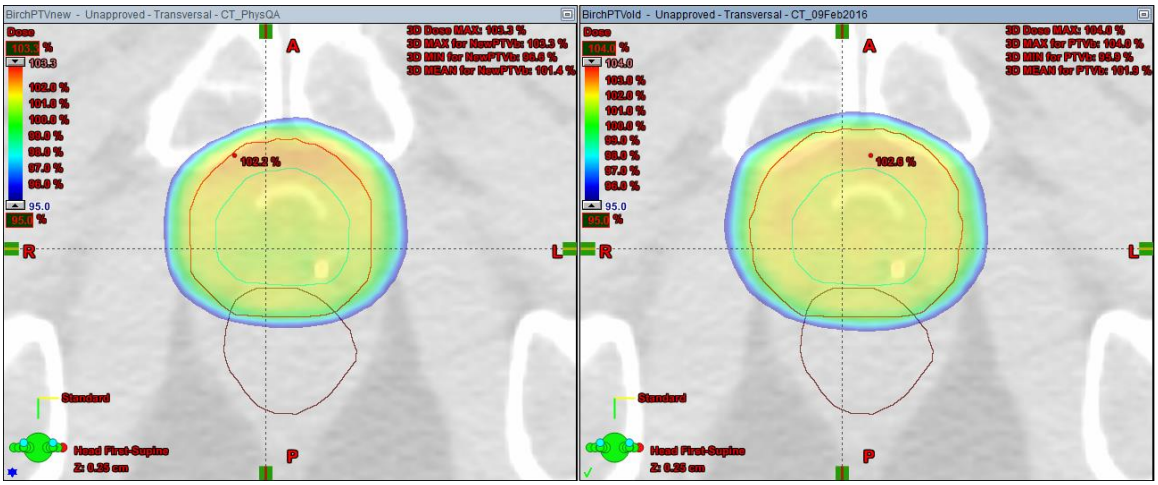

Figure 17: Comparison of original PTV and new PTV for one patient (red contours). The new PTV is on the left. The original PTV is on the right.

Two important organs-at-risk (OARs) for prostate treatment are the rectum and the bladder. Figure 18 shows that there is evidence of increased bladder and rectum sparing with the new PTV margin. Both the rectum and bladder are irradiated to a lower dose overall. Figure 18 shows that the 95% isodose infringes on these structures to a lesser degree with the new PTV compared to the original PTV.

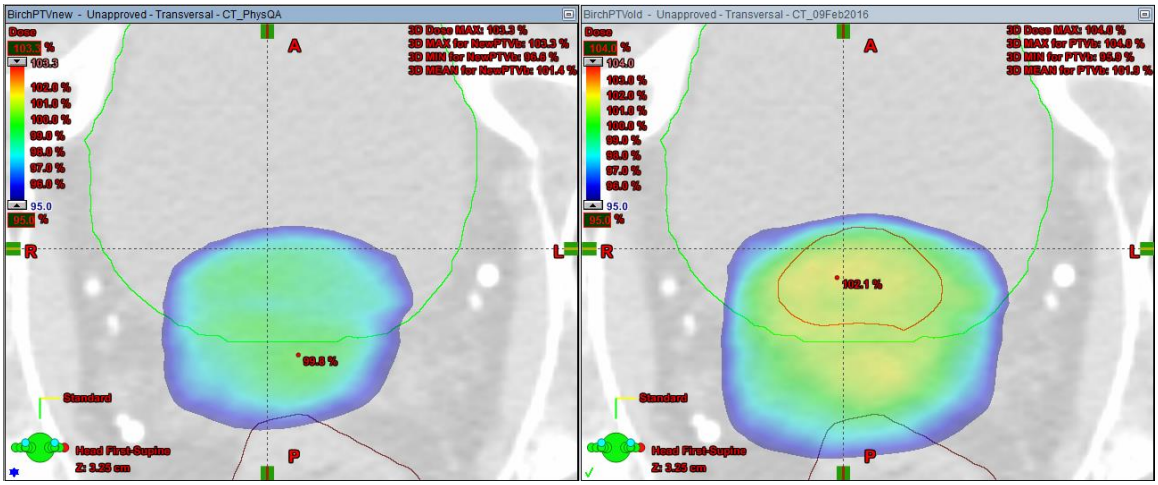

Figure 18: Dose distribution near important organs-at-risk. The bladder (green contour) and rectum (brown contour) are both shown here. The new PTV plan is on the left. The original PTV plan is on the right. The PTV (red contour) is visible in this slice on the original plan only. The PTV used for the new plan does not extend into this slice.

The normal tissue sparing becomes more apparent when looking at the dose volume histogram (DVH) (Figure 19). A smaller volume of the bladder and rectum are irradiated and to a lower dose when using the plan with the new PTV. Additional statistics are included in Table 34, which also show sparing in the bladder (green contour) and the rectum (brown contour). In addition to the bladder and rectum, the left and right femoral heads (pink and orange contours) also show evidence of sparing. Meanwhile the original and new PTVs are irradiated to the same coverage: 98% of the PTV receives 100% of the prescription dose.

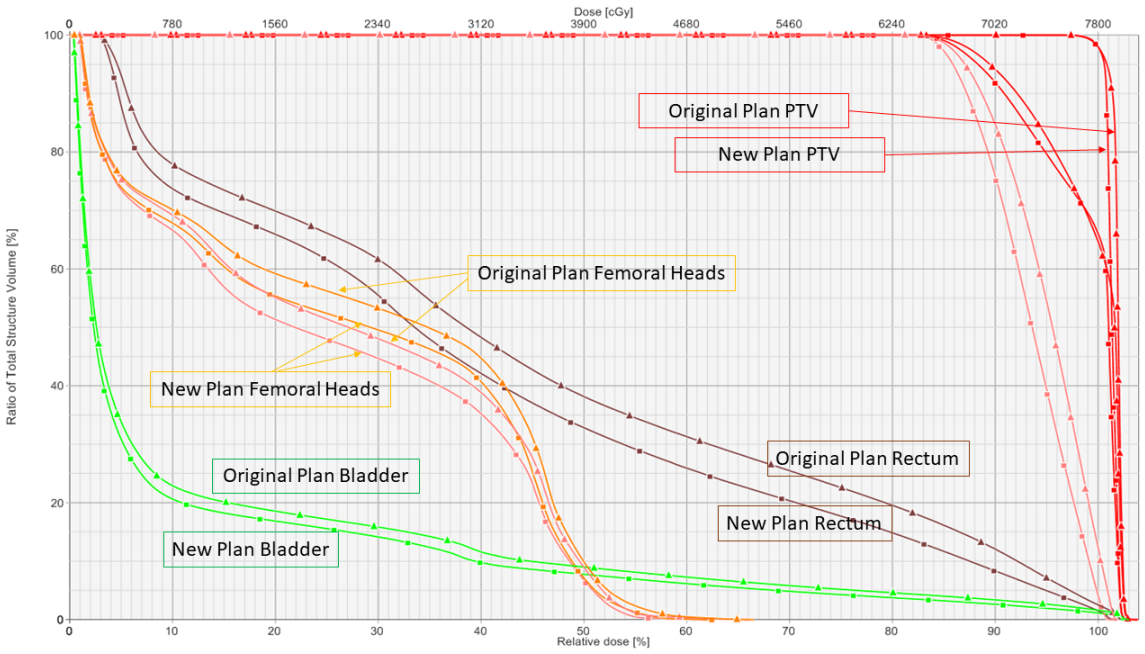

Figure 19: Dose-volume histogram (DVH) of PTV and OAR volumes under both the new and original plans. The new and original PTVs are irradiated to the same coverage. The OAR all receive lower dose when using the new PTV compared to the original PTV.

	Bladder			Rectum		
	Original PTV	New PTV	Improvement	Original PTV	New PTV	Improvement
V70 (%)	*3.4*	*2.7*	*0.7*	*12.3*	*8.5*	*3.8*
V50 (%)	*6.6*	*5.6*	*1.0*	*28.8*	*23.5*	*5.3*
V30 (%)	*12.6*	*10.5*	*2.1*	*50.1*	*43.8*	*6.3*
V20 (%)	*17.0*	*15.4*	*1.6*	*65.5*	*60.9*	*4.6*

Table 34: Relative volume receiving at least 70, 50, 30, and 20 Gy, respectively.

6 Discussion

6.1 Analysis of Patient Specific Factors for Real Patients

Collecting all PSFs posed some challenges due to accessibility to patient records and data availability. Originally, 17 PSFs were to be recorded. From these, 2 (Race and Staging) had to be omitted. Only 1 (ECOG status) of the remaining 15 PSFs did not have a value associated with each patient.

Pearson and Spearman correlation tests were performed on each of the PSFs in order to identify useful indicators related to intra-fraction prostate motion and intra- and inter-fraction prostate rotation. The correlations between PSFs and longitudinal or vertical motion were of particular interest as there was a larger deviation between the patients along these axes, allowing for a greater degree of personalization. Similarly for rotations, the pitch was considered more important than the roll and yaw axes. Overall, there were no outstandingly strong correlations between any of the PSFs and motion/rotation (Table 9 - Table 14) for n=21 patients.

As there were no strong correlations, it is difficult to say which PSFs may be important to monitor in use in predicting intra-fraction motion, but one of interest may be ECOG status. The Pearson and Spearman correlations between ECOG and mean longitudinal intra-fraction motion were 0.38 and 0.51, respectively, which is not statistically powerful. However, by breaking down the ECOG into two groups of patients, those experiencing no difficulty in their daily life (ECOG = 0) and those experiencing at least some difficulty in

their daily life (ECOG > 0) (rather than the original six groups), there does appear to be a difference in the mean longitudinal intra-fraction motion (Table 35). The sample size for the two groups is still small, but it is interesting nonetheless.

	ECOG = 0	ECOG > 0
Number of Patients	*9*	*8*
Mean Intra-fraction Lng Motion [mm]	*-0.18*	*0.83*

Table 35: Mean intra-fraction longitudinal motion between ECOG status

Another PSF of interest was BMI since a previous study had found a correlation between BMI and intra-fraction motion. Namely, a larger BMI was related to a smaller standard deviation of intra-fraction motion, in particular along the longitudinal axis [23]. Figure 20 shows the relationship found here between standard deviation of intra-fraction longitudinal motion and BMI. While there is one outlier point that happens to be associated with a large BMI and a small motion, there doesn't appear to be any underlying relationship; however, more data is needed to verify this.

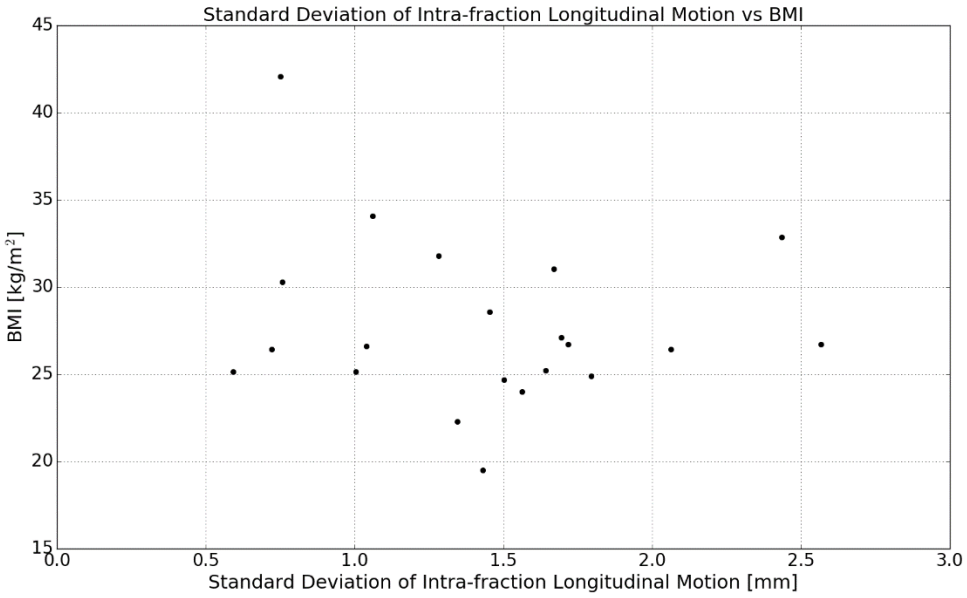

Figure 20: Scatterplot between BMI and standard deviation of intra-fraction longitudinal motion

6.2 Time Dependent Data

One of the more significant results of the study was the correlation between vertical and longitudinal intra-fraction motions. This type of correlation has been seen before by others [41]. These motions are likely related due to pelvic tilt, but also possibly rectal distension. Both these factors affect motion along the longitudinal and vertical axes where the majority of motion occurs.

A second result was the lack of correlation between intra-fraction motion and all fractions timestamps. Table 36 provides the range of times sampled for each timestamp. The magnitude of all Pearson and Spearman correlations between time stamps and intra-fraction motions were less than 0.30.

Time Stamp	N	Mean [s]	Standard Deviation [s]	Minimum [s]	Maximum [s]
First Beam	*668*	*112.7*	*53.3*	*67.0*	*676.0*
Last Beam	*663*	*280.3*	*61.8*	*197.0*	*891.0*
Total Time	*658*	*307.5*	*62.2*	*210.0*	*916.0*
Beam-on Time	*663*	*167.5*	*23.4*	*110.0*	*346.0*

Table 36: Range of sampled timestamps and correlations with intra-fraction motion

As mentioned earlier, previous studies had found a relationship between these parameters. Namely, as treatment times decrease, motion tends to decrease as well [33,38,39]. Two of these studies ([33] and [39]) used real-time tracking throughout treatment. Tracking times for treatment fractions in these studies ranged from 8.1 min to 24 min and 2.5 min to 18.4 min, respectively, and motion tended to increase throughout treatment. The third study ([38]) took multiple EPID images during each treatment beam in order to monitor the amount of intra-fraction motion. They found that a larger amount of motion occurred in approximately the first 2 minutes of treatment, after which the motion lessened.

While the time-dependence results found in this thesis work may be an accurate representation of the process at VIC, the lack of correlation compared to the results seen in the literature could also be due to poor resolution of timestamps/sampling of the motion. There were only three timestamps recorded: Time of first beam, time of last-beam, and the total time that the patient file was open. The time between the first and last beam timestamps was also inferred (called ‘Beam-on Time’), but found to be uncorrelated as well. It is possible that this time-dependence occurred too quickly to capture with our sampling, given that our mean treatment times were around 2 min to the first treatment beam, then 2-3 min to deliver the beams.

An interesting approach that could be implemented at VIC is to take an MV image perpendicular to the longitudinal and vertical axes, immediately prior to treatment and simultaneous with another time stamp, in addition to the pre-treatment and post-treatment imaging. This would provide insight into how the prostate has moved prior to the treatment, but after image guidance with on-board imaging. This new timestamp could be used to test the influence of time on intra-fraction motion by measuring the motion during the time between pre-treatment image guidance and the MV image (very close to treatment beam) and the time between the MV image and the post-treatment imaging.

More generally, this additional image would provide insight into the degree to which intra-fraction motion of the prostate should be treated as a systematic error versus random error. While intra-fraction motion will always introduce random error (motion about a mean position), the results presented in this thesis suggest there may be a not insignificant posterior shift in the mean prostate position in the population of patients caused by intra-fraction motion. It has been seen elsewhere [41,42,43] that the majority of this intra-fraction prostate shift may actually occur before the first treatment beam, possibly due to a slow relaxation of pelvic muscles after the patient has been positioned on the couch. The timing of this shift with respect to the image guidance and treatment beams (along with the reproducibility of the timing of the shift across the population of patients) is important in determining its significance as a systematic error.

Finally, with knowledge of the shift's time-dependence, treatment may be better personalized through PSFs as well. BMI and weight for example, are two PSFs that, at least

intuitively, could affect the rate of prostate shift rather than the final prostate shift. Of course, a real-time imaging system would be the best method to identify the time-dependence of prostate motion [39,44] and its possible relationship with PSFs.

6.3 Rectal Distension

The data showed a relationship between changes in gas and intra-fraction motion during treatment fractions. The mean motion was strongly affected along the longitudinal and vertical axis. Motion along these axes was expected due to the position of the rectum under the prostate. The mean lateral motion did not appear to be affect by a change in gas. The standard deviation of intra-fraction motion also did not appear to be affected by a change in gas.

The inter-fraction rotation data suggested that the presence of pre-treatment gas affected the rotation of the prostate. The pitch and roll were affected, which is again understandable given the position of the prostate with respect to the rectum. The standard deviation was affected along the pitch rotation axis to a small degree. This may be statistical or it may be due to the poor resolution of gas classification used. For example, a patient may have had a “small” amount of gas or a “large” amount of gas, but both of these cases would have been binned together as “having gas”. This would have caused a change in the standard deviation of the pitch rotation (and likely influence the other rotational axes to some degree as well). Furthermore, gas information can be challenging to infer from the images given the overlapping anatomy present in the orthogonal images. This contributes to errors (variability) in classifying the presence of gas from treatment fraction images.

The mean intra-fraction pitch rotation was more negative (i.e. rotated from the patient's head towards their feet) when a change in gas occurred during treatment fraction. The other axes were not affected intra-fractionally by a change in gas. The standard deviation of all the intra-fraction rotations were not affected in a meaningful way.

Overall, gas had the largest influence on the mean prostate shifts vertically and longitudinally, where a change in gas resulted in a prostate shift of approximately 1mm along those axes. A 1mm shift is important, but is also within the standard deviation of typical mean prostate motion along those axes over a course of radiation therapy. As a random occurrence, this gas change is therefore likely accounted for by a population-based PTV margin. However, patients who experience more frequent gas (multiple fractions) may not be adequately covered.

Table 37 breaks down patient treatment fractions in terms of a change in gas state. While a change in gas only occurred in 27 of the 669 total fractions (4.0%) surveyed here – a seemingly small number – these higher risk fractions were not uniformly distributed amongst the patients. More specifically, only 6 out of the 21 total patients (28.6%) had multiple gas change fractions (highlighted in red), but these patients accounted for 24 of the 27 total gas change fractions (88.9%). Patient 1 was the most remarkable in this regard, as a change in gas occurred in 7 of their 38 fractions (18.4%). Given the contrast between patients in terms of the number of fractions affected by gas, along with the impact of gas

on intra-fraction motion, there is possibly an opportunity here for further PTV margin personalization.

Patient	Total Fractions	Fractions with Gas Change	Percent of Fractions with Gas Change (%)
1	*38*	*7*	*18.4*
2	*39*	*3*	*7.7*
3	*28*	*0*	*0.0*
4	*39*	*0*	*0.0*
5	*28*	*0*	*0.0*
6	*28*	*0*	*0.0*
7	*28*	*4*	*14.3*
8	*28*	*0*	*0.0*
9	*39*	*3*	*7.7*
10	*39*	*0*	*0.0*
11	*27*	*0*	*0.0*
12	*28*	*0*	*0.0*
13	*26*	*3*	*11.5*
14	*28*	*0*	*0.0*
15	*39*	*0*	*0.0*
16	*38*	*1*	*2.6*
17	*28*	*4*	*14.3*
18	*37*	*1*	*2.7*
19	*28*	*0*	*0.0*
20	*28*	*0*	*0.0*
21	*28*	*1*	*3.6*

Table 37: Occurrence of a change in gas broken down by individual patient. Patients with greater than 1 fractions with a gas change are highlighted in red.

As far as the rotations are concerned, gas had the largest impact on mean intra-fraction pitch with a roughly 4° difference between fractions with and without gas change. There is not a lot of information available about the dosimetric consequences of prostate rotation. These rotations may not be consequential to the prostate [45], which is understandable considering that the prostate is roughly spherical. That being said, rotational errors may be of greater important to the seminal vesicles, as they may experience greater under-dosage due to rotation given their anatomical position [46].

6.4 Population-based Statistics

One of the more important results of the study was a systematic mean intra-fractional prostate shift posteriorly, similarly observed by others [39,40,41,44] which may be due to gradual relaxation of the pelvic muscles. Longitudinal and lateral intra-fractional shifts were less significant. The standard deviation of intra-fraction motion was greater along the vertical and longitudinal axes than the lateral axis. This is further evidence that the range of motion is greatest along the vertical and longitudinal axes.

The directionality of intra-fraction motion again shows a greater mean prostate motion posteriorly, rather than anteriorly. The anterior and posterior directions have similar standard deviations, however, which suggests that the motion is consistent in either direction. In contrast to this, the longitudinal directionality reveals a larger standard deviation in the superior motion than the inferior motion, which suggests that inferior intra-fraction motion is more consistent than superior intra-fraction motion.

The extent of inter- and intra-fraction rotations tended to be important only along the pitch rotation axis. Inter-fraction pitch had mean value of -6.4°, which is larger than in the other axes. A negative pitch could be caused by a relaxation of the pelvic muscles during treatment fractions compared to during CT simulation, possibly due to patient stress. It has been suggested that systematic pitch rotations greater in magnitude than 5° are dosimetrically important [47], but this is not well known.

Another interesting result is the directionality of inter-fraction pitch rotation. The extent of mean negative inter-fraction pitch is large (-10.0°) and approximately double the extent of mean positive inter-fraction pitch (4.8°). This result echoes the result of the mean inter-fraction motion: Prostate treatment fraction rotation anterior to inferior may be more accessible to the patient than rotation anterior to superior, at least compared to simulation CT. This may be caused by abdominal strain during simulation CT compared to subsequent treatment fractions, making pelvic rotations more accessible in one direction compared to the other. Further investigation into the dosimetric effects of the rotations are required to gauge the importance of these rotations.

6.5 Predictive Algorithm Benchmarking

Several regression algorithms were investigated for use in predicting a patient's intra-fraction motion based on their unique PSFs. While there was uncertainty in the generation of the simulated data (as it is impossible to totally mimic the real distributions of patient motion and PSFs), this benchmarking provides some groundwork for how they can be expected to perform on similar data going forward.

Based on the results of the simulations, it appears as though the Ridge model is a good candidate for future use as it performed at or near the best of all the algorithms throughout. It is a relatively simple model which uses a penalty term in the objective function to limit the influence of noise in the determination of model parameters, and is simple, easily interpretable, and robust to noise. Compared to the LASSO model (the other regularized regression model), the Ridge model also deals with collinearity between PSFs better, which is also advantageous.

Another key result of the simulations was a significant improvement in algorithm predictive performance within the first 50-100 patients. Furthermore, the algorithms experienced this improvement with different levels of noise. If the performance of the algorithms does not improve after using approximately 100 patients for training, it would seem unlikely that predicting intra-fraction motion based on PSFs is useful/feasible.

Finally, a large sample of real patient data is required to understand the true predictive capabilities of any of the algorithms because it is impossible to know the underlying correlations between PSFs and motion given the noise in the collected motion data. The performance of the algorithms, even with a large sample of patients, is limited by the underlying noise in the dataset.

6.6 Predictive Algorithms for Real Patient Data

The LASSO model performed the best of all the predictive algorithms on the patient data. It predicted mean intra-fraction motion to within 0.8mm MAE and standard deviation of

intra-fraction to within 0.4mm. Unfortunately, this algorithm also ignored the input of the PSFs to the largest degree. It effectively predicted an identical motion for each patient making it a robust predictor, which is important, but not necessarily a personalized predictor. Ultimately, the collection of a larger quantity of data will determine the feasibility of this kind of personalization.

The collection of more data will allow users to better select which PSFs are relevant to predicting intra-fraction motion. One simple approach for doing this is to threshold the PSFs based on their correlation. For example, any PSF with a Pearson correlation greater than 0.4 could be included as an input to the algorithm, whereas PSFs with a Pearson correlation below 0.4 would be excluded. Another approach is to choose a fixed number of inputs and populate those inputs based on the performance of the algorithm. For example, 5 input parameters may be chosen for the algorithm and the performance of the algorithm could be evaluated by cross-validation using only the 5 most correlated PSFs in each cross-validation subset.

Parameter tuning may also be useful in identifying pertinent PSFs. In this work, the LASSO model performed better when no consideration was given to the PSFs. However, if more patients are collected, parameter tuning may reveal that the best performance comes at some value of α where some of the PSF coefficients are non-zero. This would again be useful in choosing appropriate PSFs as inputs to the algorithm.

6.7 PTV Margin Evaluation

Two approaches to PTV margin generation were presented in this work: A population-based approach, and a personalized approach.

The population-based approach revealed a possible opportunity to reduce margins for prostate external beam radiation therapy from the 7mm which is currently used at VIC. This work would see treatment planning PTV margin (rounded up to the nearest integer) of 5mm, 6mm, and 4mm in each of the vertical, longitudinal, and lateral directions. This corresponds to a reduction of 2mm vertically, 1mm longitudinally, and 3mm laterally. The results point to a lateral PTV margin reduction especially, considering that both the mean and standard deviation of intra-fraction motion is roughly halved along the lateral axis compared to the longitudinal and vertical axes.

Personalized margin methodology was similar to population-based margin generation, except the population random motion error was replaced by a personalized random error. For n=21 patients, however, the personalized margin was identical to the population-based margin, since the best algorithm did not rely on PSF data.

The new margin was tested on one patient and appropriately covered the PTV with a uniform dose distribution between 95% and 107% of the prescribed dose. The new margin also showed evidence of normal tissue sparing, based on both the dose-volume histogram information and the dose distributions. While these results only reflect the treatment

planning dose distributions, it can be concluded that there is potential clinical benefit to normal tissues by reducing PTV margins.

Moving forward, the greatest opportunity for personalization appears to be by accounting for inter-patient differences in mean intra-fraction motion, particularly along the longitudinal and vertical axes (Table 38). The standard deviation in the individual patient random motion (σ_i) along any of the directions is relatively small compared to the standard deviation in the individual patient systematic motion (μ_i).

Reproducibility of 3D intra-fraction motion			
	Longitudinal	**Vertical**	**Lateral**
Standard deviation of μ_i [mm]	*0.94*	*1.04*	*0.54*
Standard Deviation of σ_i [mm]	*0.54*	*0.46*	*0.30*

Table 38: Reproducibility of 3D intra-fraction motion

Accounting for a systematic intra-fraction motion may be challenging in practice. One approach may be to identify high risk patients based on the extent of their anticipated systematic intra-fraction motion, and then put additional efforts into supressing and/or otherwise accounting for their motion. In particular, if the expected systematic motion is caused by prostate drift prior to the first treatment beam, this correction could be as simple as allowing a few extra minutes for the high-risk patient to lay on the couch prior to image guidance.

7. Conclusion

The purpose of this thesis was to investigate the feasibility of personalized PTV margins for EBRT of the prostate, with particular emphasis on regression-based predictive algorithms. Patient motions and rotations were analysed for correlation with their potential predictors, the PSFs, using Pearson and Spearman correlation tests. Population-based statistics were used to gauge the typical range of motions/rotations for prostate cancer patients at VIC, as well as identify important motion/rotation directions for personalization. Predictive algorithms were benchmarked for predicting intra-fraction motion of the prostate using clinically relevant, simulated PSFs and motion data. Benchmarked algorithms were applied to actual patient data and evaluated for use in predicting intra-fraction motion along all three spatial directions. The algorithm with the best performance (smallest error) was used to estimate a personalized PTV margin.

Overall, the results of the Pearson and Spearman correlation tests revealed no outstanding, strong correlations between any of the PSFs and motion/rotation. Motion along the vertical and longitudinal axes, as well as pitch rotation offered the largest opportunity for personalization given the relatively large standard deviations along those axes. Predictive algorithms benchmarking indicated that the Ridge regression algorithm is a good candidate for future use as it performed at or near the best of all the algorithms throughout and algorithm performance should improve significantly within the first 50-100 patients. The LASSO model performed the best on actual patient data by ignoring the input of the PSFs suggesting that a regression-based PTV margin personalization may not be useful. More

data (at least 50 patients) is required to evaluate this assertion however, as seen in the simulation results.

A population-based margin and a personalized margin were developed. The two margins were ultimately identical since the best algorithm did not rely on the PSF data. The new margin was non-isotropic and extended 5mm, 6mm, and 4mm in each of the vertical, longitudinal, and lateral directions, corresponding to a reduction of 2mm vertically, 1mm longitudinally, and 3mm laterally from the isotropic 7mm margin which is currently used at VIC. The results point to a lateral PTV margin reduction especially, considering that both the mean and standard deviation of intra-fraction motion is roughly halved along the lateral axis compared to the longitudinal and vertical axes. The new margin achieved similar PTV coverage compared to the current PTV margin, but also showed evidence of normal tissue sparing.

Beyond the primary thesis objective, other interesting results were found. A strong correlation was found between the vertical and longitudinal axes, likely the result of intra-fraction pelvic tilt and/or rectal distension. There was no correlation found between BMI and any intra-fraction motion/rotation, in contrast to results found in a previous study [23]. The time dependent data also revealed no correlation between the recorded treatment time stamps and intra-fraction motion, also in contrast to prior studies [33]. This result may be due to poor timestamp resolution, however, and could benefit by taking an additional image immediately prior to the first treatment beam. Rectal distension had a large impact (~1mm) on the mean intra-fraction motion of the prostate vertically and longitudinally. While a

single case of rectal distension over the course of approximately 35 fractions is likely acceptable using a population-based PTV margin, repeated instances of rectal distension may be significant to treatment outcome and offer opportunity for PTV margin personalization by expanding the margin of those patients most affected by rectal distension and for those undergoing hypo-fractionated (~5 fraction) treatment schedules. Population-based statistics revealed a systematic mean intra-fraction prostate shift in the posterior direction. This has been seen elsewhere [39,40,41,44] and may be due to a relaxation of the pelvic muscles. Going forward, treatment could be personalized by identifying high risk patients (possibly via PSFs) and simply allowing them more on-couch time for muscle relaxation prior to image guidance.

To the writer's knowledge, this is the first effort at generating personalized PTV margins using an all-encompassing approach via PSFs. Others have investigated the influence of a single PSF [23], or generated personalized margins based on real-time prostate motion alone [48] without consideration for motion causation, however.

Moving forward it will be important to investigate whether there is merit to personalization given the additional uncertainties introduced by using a statistical algorithm. It will likely depend on the ultimate strength of the motion predictors (PSFs) as well as the opportunity for improvement. For example, based on the results presented in this thesis, it seems much more possible that PTV margin personalization will be used along the longitudinal and vertical axes compared to the lateral axis, based solely on the inter-patient variance in motion along those axes.

7.1 Future Work

For the immediate future, a useful study at VIC is to test the time-dependence of intra-fraction prostate motion. This can be done using two methods:

1. Set aside a group of patients and allow them more on-couch pre-image guidance time for possible pelvic muscle relaxation and compare the distribution of motions/rotations with respect to patients that did not receive additional on-couch time.
2. Take an additional MV image immediately prior to the first treatment beam in order to gauge the amount of intra-fraction motion that occurs after image guidance, but before the first treatment beam with the total intra-fraction motion.

In the longer term future, moving forward toward personalized PTV margins for prostate cancer (and personalized medicine in general), it is important to amalgamate data from several institutions in order to achieve the most conclusive information. This is challenging however, not just because of patient data confidentiality and sharing of that information, but also because the process at each institution is different and therefore not necessarily representative of same information. It is still a useful endeavour as long as the patients are properly labelled with their corresponding institution, allowing for the distribution of patient attributes (PSFs, treatment parameters, motion, etc.) at each institution can be compared and contrasted.

Ideally this data collection is a fully automated process where patient data (demographic, treatment, motion, institution, etc.) are all collected and automatically synced into a common database without middle-people such as physicians and therapists. One could simply select the type of treatment (prostate EBRT) and the motion collection would be automated through post-treatment imaging and matching, for example, and uploaded to the database (after proper anonymization).

The ideal motion tracking would be performed using real-time tracking software that is currently available (such as the Calypso 4D Localization System). This would allow for statistics to be performed on a single fraction basis, offering better insight into the day-to-day differences in patient anatomy, physiology, and behaviour, and their resultant impact on intra-fraction motion of the prostate. In particular, real-time motion tracking would provide crucial information on the time dependence of prostate motion. A patient may have a stable prostate throughout the treatment fraction, a slow prostate drift during the fraction, or a more erratic prostate motion. Time-dependent prostate motions such as these have been seen before [44]. Furthermore, patients may be able to be classified according to their respective type of prostate motion based on their unique PSFs. Patients classified with erratic prostate motion could then be planned with a larger PTV compared to patients classified with a stable prostate, for example.

Finally, given the decreasing cost of data storage and genome sequencing techniques [49], it is not inconceivable that one day a patient's genome could be sequenced as a regular part of treatment planning. This could provide the ultimate predictors (PSFs) and could be

identified using algorithms discussed in this thesis. This type of data could be also useful for personalized prescription radiation doses via inter-patient differences in radio sensitivity for example, in addition to intra-fraction motion predictions.

Bibliography

[1] R. Baskar, K. A. Lee, R. Yeo and K. W. Yeoh, "Cancer and Radiation Therapy: Current Advances and Future Directions," *International Journal of Medical Sciences,* vol. 9, no. 3, pp. 193-199, 2012.

[2] F. H. Attix, Introduction to Radiological Physics and Radiation Dosimetry, Wiley-Interscience, 1986.

[3] H. C. John and J. R. Cunningham, The Physics of Radiology, 4th ed., Charles C. Thomas, 1983.

[4] M. Joiner and A. van der Kogel, Basic Clinical Radiobiology, Boca Raton, FL: CRC Press, 2009.

[5] ICRU Report 62: Prescribing, Recording, and Reporting Photon Beam Therapy (Supplement to ICRU Report 50), Bethesda, Maryland: International Commission on Radiation Units and Measurements.

[6] M. van Herk, "Errors and margins in radiotherapy," *Seminars in Radiation Oncology,* vol. 34, no. 10, pp. 52-64, 2004.

[7] J. A. Antolak and I. I. Rosen, "Planning target volumes for radiotherapy: how much margin is needed?," *International Journal of Radiation Oncology Biology Physics,* vol. 44, no. 5, pp. 1165-1170, 1999.

[8] T. Hastie, R. Tibshirani and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Second Edition, New York, New York: Springer, 2009.

[9] E. B. Poggorsak, Radiation Physics for Medical Physics, Springer-Verlag Berlin Heidelberg, 2006.

[10] S. H. Levitt, J. A. Purdy, C. A. Perez and P. Poortmans, Technical Basis of Radiation Therapy - 5th Edition, Springer-Verlag Berlin Heidelberg, 2012.

[11] F. M. Khan, The Physics of Radiation Therapy - Edition 5, Philadelphia, PA: LIPPINCOTT WILLIAMS & WILKINS, 2014.

[12] M. Murphy, J. Balter, Balter S, J. J. BenComo, I. J. Das, S. B. Jiang, C. M. Ma, G. H. Olivera, R. F. Rodebaugh, K. J. Ruchala, H. Shirato and F. Yin, "The management of imaging doses during image-guided radiotherapy: Report of the AAPM Task Group 75," *Medical Physics,* vol. 34, no. 10, pp. 4041-4063, 2007.

[13] J. Han and M. Kamber, Data Mining: Concepts and Techniques, San Francisco, California: Morgan Kaufmann Publishers, 2006.

[14] G. James, "Variance and Bias for General Loss Function," *Machine Learning,* vol. 51, pp. 115-135, 2003.

[15] "Prostate-specific antigen (PSA) test," Canadian Cancer Society, 2017. [Online]. Available: http://www.cancer.ca/en/cancer-information/diagnosis-and-treatment/tests-and-procedures/prostate-specific-antigen-psa/. [Accessed 29 June 2017].

[16] G. Welch, L. M. Schwartz and S. Woloshin, "Prostate-Specific Antigen Levels in the United States: Implications of Various Definitions for Abnormal," *Journal of the National Cancer Institute,* vol. 97, no. 15, pp. 1132-1137, 2005.

[17] "Grading prostate cancer," Canadian Cancer Society, 2017. [Online]. Available: http://www.cancer.ca/en/cancer-information/cancer-type/prostate/grading/. [Accessed 29 June 2017].

[18] A. El-Hakim and S. Moussa, "CUA guidelines on prostate biopsy methodology," *Canadian Urological Association Journal,* vol. 4, no. 2, pp. 89-94, 2010.

[19] "TNM Staging," Canadian Cancer Society, 2017. [Online]. Available: http://www.cancer.ca/en/cancer-information/diagnosis-and-treatment/staging-and-grading/staging/tnm-staging/. [Accessed 29 June 2017].

[20] "Am I at risk of prostate cancer?," Prostate Cancer Canada, 2017. [Online]. Available: http://www.prostatecancer.ca/Prostate-Cancer/About-Prostate-Cancer/Risk-Factors. [Accessed 29 June 2017].

[21] "Mean height, weight, body mass index (BMI) and prevalence of obesity, by collection method and sex, household population aged 18 to 79, Canada, 2008, 2007, to 2009, and 2005," Statistics Canada, 2015. [Online]. Available: http://www.statcan.gc.ca/pub/82-003-x/2011003/article/11533/tbl/tbl1-eng.htm. [Accessed 29 June 2017].

[22] F. Nuttall, "Body Mass Index," *Nutrition Today,* vol. 50, no. 3, pp. 117-128, 2015.

[23] A. L. Thompson, S. Gill, J. Thomas, T. Kron, C. Fox, A. Herschtal, K. H. Tai and F. Foroudi, "In pursuit of individualized margins for prostate cancer patients undergoing image-guided radiotherapy: The effect of body mass index on intrafraction prostate motion," *Clinical Oncology`,* vol. 23, pp. 449-453, 2011.

[24] M. Ng, E. Brown, A. Williams, M. Chao, N. Lawrentschuk and R. Chee, "Fiducial markers and spacers in prostate radiotherapy: current applications," *BJU International,* vol. 113, pp. 13-20, 2014.

[25] "ECOG Performance Status," ECOG-ACRIN cancer research group, 2016. [Online]. Available: http://ecog-acrin.org/resources/ecog-performance-status. [Accessed 29 June 2017].

[26] "Types of Diabetes," Diabetes Canada, 2017. [Online]. Available: http://www.diabetes.ca/about-diabetes/types-of-diabetes. [Accessed 29 June 2017].

[27] M. N. Piero, G. M. Nzaro and J. M. Njagi, "Diabetes mellitus - a devastating metabolic disorder," *Asian Journal of Biomedical and Pharmaceutical Sciences,* vol. 4, no. 40, pp. 1-7, 2014.

[28] "IBS Overview," GI Society Canadian Society of Intestinal Research, 2017. [Online]. Available: http://www.badgut.org/information-centre/a-z-digestive-topics/ibs/. [Accessed 29 June 2017].

[29] K. Occhipinti and J. W. Smith, "Irritable Bowel Syndrome: A Review and Update," *Clinics in Colon and Rectal Surgery,* vol. 25, pp. 46-52, 2012.

[30] L. Saha, "Irritable bowel syndrome: Pathogenesis, diagnosis, treatment, and evidence-based medicine," *World Journal of Gastroenterology,* vol. 20, no. 22, pp. 6759-6773, 2014.

[31] "Chronic obstructive pulmonary disease (COPD)," Wolrd Health Organization, November 2016. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs315/en/. [Accessed 29 June 2017].

[32] "Prostate Cancer Risk Factors," American Cancer Society, 16 February 2016. [Online]. Available: https://www.cancer.org/cancer/prostate-cancer/causes-risks-prevention/risk-factors.html. [Accessed 29 June 2017].

[33] J. Sheng Li, M. H. Lin, M. K. Buyyounouski, E. M. Horwitz and C. M. Ma, "Reduction of prostate intrafractional motion from shortening the treatment time," *Physics in Medicine and Biology,* vol. 58, no. 14, pp. 4921-4932, 2013.

[34] I. Ogino, H. Uemura, T. Inoue, Y. Kubota, K. Nomura and N. Okamoto, "Reduction of prostate motion by removal of gas in rectum during radiotherapy," *International Journal of Radiation Oncology and Biology Physics,* vol. 72, no. 2, pp. 456-466, 2008.

[35] M. Mohri, A. Rostamizadeh and A. Talwalker, Foundations of Machine Learning, Cambridge: The MIT Press, 2012.

[36] A. Smola and B. Scholkopf, "A tutorial on Support Vector Regression," *Statistics and Computing,* vol. 14, pp. 199-222, 2004.

[37] K. Ullman, H. Ning, R. Susil, A. Ayele, L. Jocelyn, J. Havelos, P. Guion, H. Xie, L. Guang, B. C. Arora, A. Cannon, R. W. Miller, C. N. Coleman, K. Camphausen and C. Menard, "Intra- and inter-radiation therapist reproducibility of daily isocenter verification using prostate fiducial markers," *Radiation Oncology,* vol. 1, no. 2, 2006.

[38] T. M. Haska, H. Honore, L. P. Muren, M. Hoyer and P. R. Poulsen, "Intrafraction changes of prostate prostate position and geometrical errors studied by continuous electronic portal imaging," *Acta Oncologia,* vol. 47, no. 7, pp. 1351-1357, 2008.

[39] K. Langen, T. R. Willoughby, S. L. Meeks, A. Santhanam, A. Cunningham, L. Levine and P. Kupelian, "Observations on real-time prostate gland motion using electromagnetic tracking," *International Journal of Radiation Oncology Biology Physics,* vol. 71, no. 4, pp. 1084-1090, 2008.

[40] P. Poulsen, L. P. Muren and M. Hoyer, "Residual set-up errors and margins in on-line image-guided prostate localization in radiotherapy," *Radiotherapy and Oncology,* vol. 85, pp. 201-206, 2007.

[41] T. Budiharto, P. Slagmolen, K. Haustermans, F. Maes, S. Junius, J. Verstraete, R. Oyen, J. Hermans and F. Van den Heuvel, "Intrafractional prostate motion during online image guided intensity-modulated radiotherapy for prostate cancer," *Radiotherapy and Oncology,* vol. 98, pp. 181-186, 2011.

[42] J. Adamson and Q. Wu, "Inferences about prostate intrafraction motion from pre- and posttreatment volumetric imaging," *International Journal of Radiation Oncology and Biology Physics,* vol. 75, no. 1, pp. 260-7, 2009.

[43] T. Kron, J. Thomas, C. Fox, A. Thompson, R. Owen, A. Herschetal, A. Haworth, K. H. Tai and F. Foroudi, "Intra-fraction prostate displacement in radiotherapy estimated from pre- and post-treatment imaging of patients with implanted fiducial markers," *Radiotherapy and Oncology,* vol. 95, pp. 191-197, 2010.

[44] Y. Xie, D. Djajaputra, C. R. King, S. Hossain, L. Ma and L. Xing, "Intrafractional motion of the prostate during hypofractionated radiotherapy," *International Journal of Radiation Oncology Biology Physics,* vol. 72, no. 1, pp. 236-246, 2008.

[45] Y. van Herten, J. B. van de Kamer, N. van Wieringen, B. R. Pieters and A. Bel, "Dosimetric evaluation of prostate rotations and their correction by couch rotations," *Radiotherapy and Oncology,* vol. 88, pp. 156-162, 2008.

[46] I. Lips, U. A. van der Heide, A. N. Kotte, M. van Vulpen and A. Bel, "Effect of translational and rotational errors on complex dose distributions with off-line and on-line position verification," *International Journal of Radiation Oncology Biology Physics,* vol. 74, no. 5, pp. 1600-1608, 2009.

[47] G. Cranmer-Sargison, "A treatment planning investigation into the dosimetric effects of systematic prostate patient rotational set-up errors," *Medical Dosimetry,* vol. 33, no. 3, pp. 199-205, 2008.

[48] P. Rassiah-Szegedi, B. Wang, M. Szegedi, J. Tward, H. Zhao, Y. J. Huang, V. Sarkar, D. Shrieve and B. Salter, "Inividualized margins for prostate patients using a wireless localization and tracking system," *Journal of Applied Clinical Medical Physics,* vol. 12, no. 3, pp. 194-204, 2011.

[49] "Illumina Promises To Sequence Human Genome For $100 -- But Not Quite Yet," Forbes, 9 January 2017. [Online]. Available: https://www.forbes.com/sites/matthewherper/2017/01/09/illumina-promises-to-sequence-human-genome-for-100-but-not-quite-yet/#33d3beb8386d. [Accessed 29 June 2017].

Appendix

Example csv file templates and assorted plots with error bars are presented here.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	date	time	patient_n	fraction_n	pre_vrt	pre_lng	pre_lat	post_vrt	post_lng	post_lat	pre_gas	post_gas	time_first	time_last	time_total
2	27-Jul-16	13:21	12	1	0	-0.4	-0.3	-0.09	0.05	0.17	N	N	1:28	5:41	6:07
3	28-Jul-16	13:20	12	2	0	-0.1	-0.3	-0.17	0.16	0.07	N	N	1:33	4:19	4:45
4	29-Jul-16	10:51	12	3	-0.1	-0.6	-0.6	-0.27	0.28	0.05	N	N	1:50	4:58	5:23
5	02-Aug-16	10:44	12	4	0.7	-1.2	-0.4	n/a	n/a	n/a	N	N	3:03	5:49	6:21
6	03-Aug-16	11:18	12	5	-0.2	-0.3	-0.4	-0.36	0.55	0.05	N	N	1:33	4:22	4:47
7	04-Aug-16	12:56	12	6	-0.4	-0.3	-0.4	n/a	n/a	n/a	N	N	1:54	4:43	5:09
8	05-Aug-16	9:09	12	7	0.1	-0.3	-0.4	n/a	n/a	n/a	N	N	2:46	5:32	5:58
9	08-Aug-16	9:20	12	8	0.1	-0.7	-0.7	-0.04	0.08	0.05	N	N	2:06	4:56	5:24
10	09-Aug-16	8:33	12	9	0.2	-0.9	-0.5	-0.03	0.08	0.05	N	N	1:41	4:51	5:19
11	10-Aug-16	11:02	12	10	0.3	-0.8	-0.8	n/a	n/a	n/a	N	N	2:09	4:59	5:29
12	11-Aug-16	8:40	12	11	-0.5	-0.6	-0.4	-0.2	0.29	0.03	Y	Y	2:19	5:08	5:34
13	12-Aug-16	8:42	12	12	0.2	-0.9	-0.2	-0.31	0.2	0.15	N	N	1:54	4:42	5:11
14	15-Aug-16	8:25	12	13	0.4	-0.9	-0.3	-0.16	0.2	0.03	N	N	3:02	6:12	6:38
15	16-Aug-16	10:30	12	14	0.4	-0.5	-0.1	-0.42	0.24	0.1	Y	Y	1:34	4:23	4:51
16	17-Aug-16	8:11	12	15	-0.5	-0.3	-0.5	n/a	n/a	n/a	N	N	1:49	4:40	5:13
17	18-Aug-16	8:40	12	16	0.5	-1.1	-0.2	-0.61	0.53	0.07	N	N	1:35	4:21	4:47
18	22-Aug-16	8:32	12	17	0.6	-0.9	-0.4	n/a	n/a	n/a	N	n/a	1:55	n/a	n/a
19	23-Aug-16	15:27	12	18	-0.2	-0.5	-0.3	-0.44	0.51	0	N	N	1:49	4:35	5:01
20	24-Aug-16	9:18	12	19	0.1	-0.6	-0.4	-0.37	0.3	0.05	N	N	1:32	4:22	4:52
21	25-Aug-16	11:46	12	20	0	-0.6	-0.2	-0.5	0.62	0.06	N	N	1:29	4:15	4:42
22	29-Aug-16	11:42	12	21	0	-0.4	-0.5	-0.5	0.6	0.04	N	N	1:29	4:39	5:08
23	30-Aug-16	10:58	12	22	0.6	-0.9	-0.3	-0.73	0.78	0.2	N	N	1:32	4:19	4:45
24	31-Aug-16	13:45	12	23	-0.2	-0.1	-0.6	-0.18	0.18	0.03	N	N	1:24	4:12	4:38
25	01-Sep-16	9:02	12	24	-0.4	-0.4	-0.4	-0.25	0.21	0.09	N	N	2:04	6:15	6:41
26	02-Sep-16	11:47	12	25	0.3	-1	-0.4	-0.79	0.84	0.02	N	N	1:33	4:19	4:44
27	06-Sep-16	9:04	12	26	-0.1	-0.4	-0.5	-0.33	0.18	0.01	Y	Y	2:17	5:06	5:35
28	07-Sep-16	10:24	12	27	-0.5	-0.1	-0.1	0.04	-0.16	0.02	N	N	1:50	4:40	5:07

Figure 21: Example of time dependent csv file input for one patient.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	patient_n	psa_score	primary_g	secondary	total_glea	coresPlus	coresTota	Staging	Age	Weight	Race	Diabetes	IBS	COPD	Implants	ECOG_stat	BMI	Height	IGRT
2	12	0.83	4	4	8	3	13	T2a	78	82.7	n/a	N	N	N	N	1	26.69809	176	4
3																			
4																			
5																			
6																			
7																			
8																			

Figure 22: Example of time independent csv file input for one patient

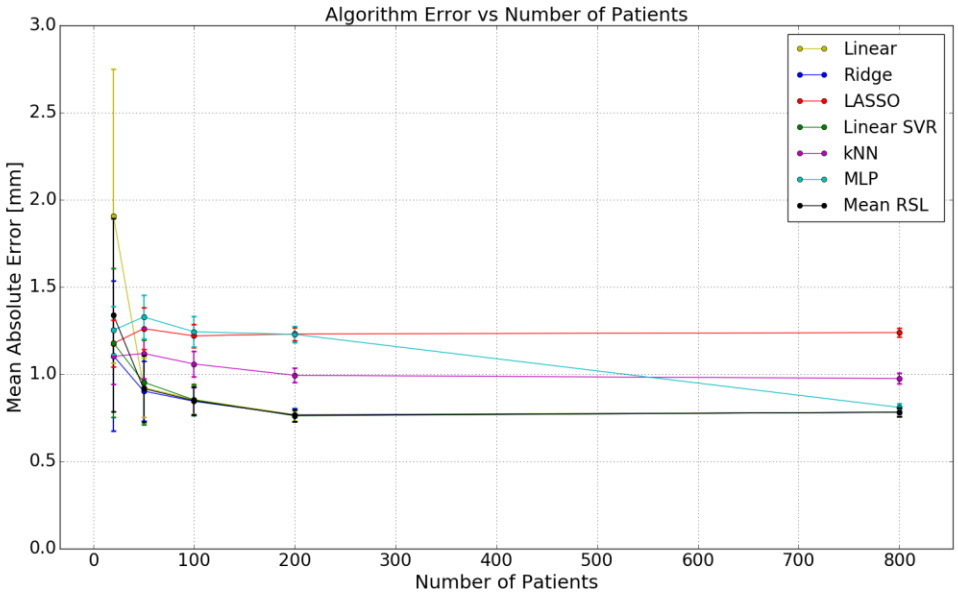

Figure 23: Projected performance of algorithms based on simulated data (with error bars)

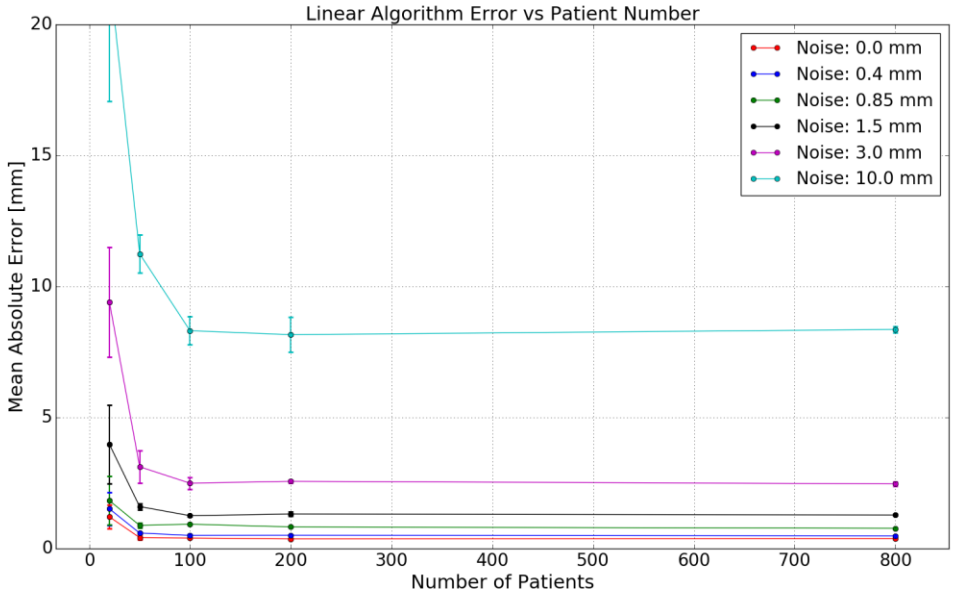

Figure 24: Effect of noise on linear model (with error bars)

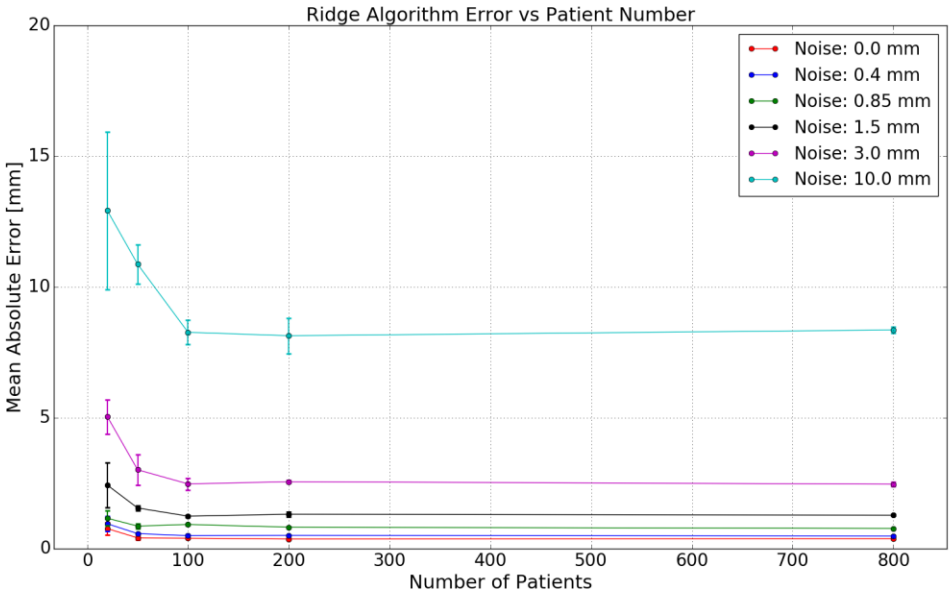

Figure 25: Effect of noise on ridge model (with error bars)

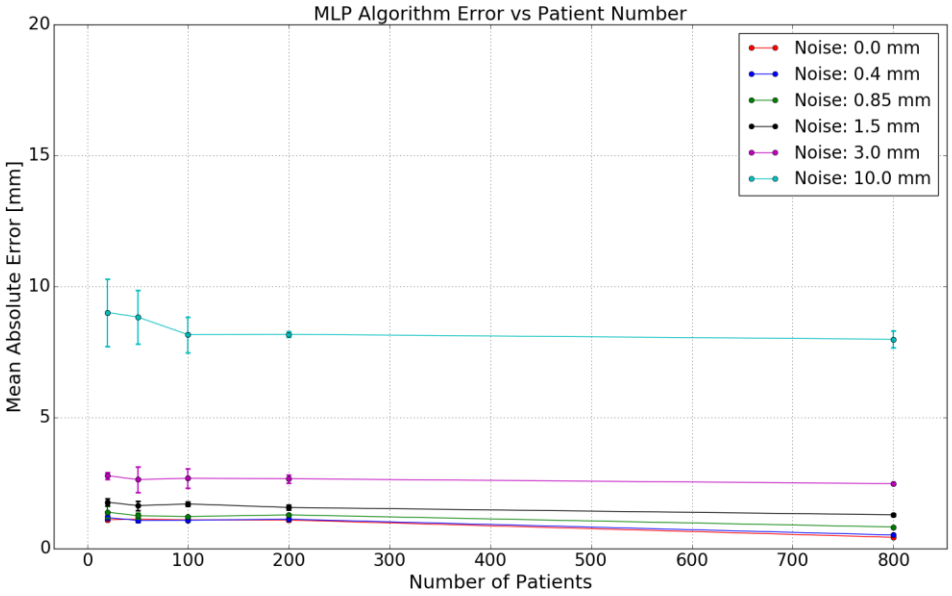

Figure 26: Effect of noise on MLP model (with error bars)