Utilizing Positron Emission Tomography Information in Lung Cancer Treatment

by

Heyse Li

A thesis submitted in conformity with the requirements
for the degree of Masters of Applied Science
Graduate Department of Mechanical and Industrial Engineering
University of Toronto

Abstract

Utilizing Positron Emission Tomography Information in Lung Cancer Treatment

Heyse Li
Masters of Applied Science
Graduate Department of Mechanical and Industrial Engineering
University of Toronto
2013

We explore both robust biologically guided intensity-modulated radiation therapy (BG-IMRT) and pattern recognition to identify responders to cancer treatment for lung cancer. Heterogeneous dose prescriptions that are derived from biological images are subject to uncertainty, due to potential noise in the image. We develop a robust optimization model to design BG-IMRT plans that are de-sensitized to uncertainty. Computational results show improvements in tumor control probability and deviation from prescription dose compared to a non-robust model, while maintaining tissue dose below toxicity levels. We applied machine learning algorithms to 4D gated positron emission tomography/computed tomography (PET/CT) scans. We identified classifiers which could outperform a naive classifier. Our work shows the potential of using machine learning algorithms to predict patient response. This could hopefully lead to more adaptive treatment plans, where the clinician would adapt the treatment based on the prediction provided at certain time intervals in the treatment.

Acknowledgements

I would like to thank all the people that supported me throughout grad school. It has been said that we stand on the shoulders of giants. I'd like to think this is true not only in research but of life in general. I couldn't have done this without all of you.

To my advisor, Timothy Chan, thanks for all your guidance and support throughout undergrad and also grad school. I really learned so much from you over the past years.

To my collaborators, Jean-Pierre Bissonnette, Tom Purdie, and Nathan Becker, thanks for guiding me through your foreign world of physics and biology and putting up with all my questions.

To my labmates, thanks for your support and advice you've given me over the past couple of years. I'll know a lot of doctors in a few years.

To my classmates, thanks for helping me pass my courses through those long nights working on assignments.

To my family, thanks for keeping me alive and always being there for me.

To my friends, thanks for all the good times and fun. I would have gone crazy without you all a long time ago.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

Lung cancer

Lung cancer is the leading cause of cancer death worldwide. It is estimated that in 2008, there were 1.6 million new cases of lung cancer and 1.4 million deaths due to lung cancer (Jemal et al., 2011). Radiotherapy is a commonly used medical technique in the treatment of lung cancer (Bowen et al., 2012), where a dose of radiation is delivered through the patient to the tumor in order to kill the cancerous cells. Since the radiation damages the deoxyribonucleic acid (DNA) of both the cancerous and healthy cells (Ehrgott et al., 2008), the goal of radiotherapy is to deliver as much dose to the tumor as possible while still sparing the healthy tissue.

Intensity-modulated radiation therapy

Intensity-modulated radiation therapy (IMRT) is a technology that allows the dose of radiation to be delivered in a non-uniform pattern. IMRT utilizes a multileaf collimator, which is a collection of slabs that can block out different parts of the radiation beam allowing it discretize a large beam of radiation into many smaller beamlets. Therefore, IMRT can modulate the intensity of the beamlets at a given point resulting in the ability to deliver more complex treatment plans than previous radiotherapy delivery systems (Ehrgott et al., 2008; Romeijn and Dempsey, 2008). By conforming closer to the patient's anatomical configuration, IMRT has the potential to improve treatment quality (Romeijn and Dempsey, 2008).

Treatment plans are designed by adjusting the intensity (or weights) of the beamlets to obtain a clinically desired dose distribution. This is also known as the fluence map optimization problem. The appropriate beamlet weights can be obtained from either forward or inverse planning techniques. Forward planning is performed by adjusting the beamlet weights manually until the dose distribution is clinically acceptable. Inverse planning relies on a set of established clinical criteria and calculates the optimal beamlet intensities required to reach these criteria using optimization algorithms. The inverse planning approach to the fluence map problem can be modelled and solved using mathematical programming (Romeijn and Dempsey, 2008).

Computed tomography

Radiotherapy treatment planning begins by acquiring a computed tomography (CT) scan of the patient, which provides geometric information about the tumor and surrounding organs-at-risk (OARs). The CT scan discretizes the patient into 3-dimensional blocks of finite resolution called voxels. Treatment planning solely using CT imaging implicitly assumes that the tumor is biologically homogeneous and aims to deliver a uniform dose (Vanderstraeten et al., 2006a) as this results in the optimal tumor control probability (TCP) (Webb and Nahum, 1999; Levin-Plotnik and Hamilton, 2004). However it is known that tumors are not in general biologically homogeneous (Yang and Xing, 2005) and recent developments in imaging technology have given rise to scanners that can provide information about the heterogeneous biological activity of tumors and organs (MacManus et al., 2009).

Positron emission tomography

In contrast to CT imaging, positron emission tomography (PET) can provide information about the internal biological activity of the patient. The imaging procedure begins by injecting a small amount of a specified radio-labelled tracer into the patient which then accumulates inside cells undergoing some biological process (Gambhir, 2002) . The radioisotope emits a positron, which then collides with an electron in the tissue and this emits 2 photons. The photons are then detected by the camera (Berger, 2003). Different tracers measure different types of biological activity. Examples of activity that can be measured include: hypoxia, metabolism, proliferation, and perfusion (Gambhir, 2002).

Fluorodeoxyglucose (FDG) is the most commonly used tracer in PET imaging (MacManus et al., 2009; Bentzen and Grégoire, 2011; Bowen et al., 2012) and is usually used clinically for diagnosis and staging. FDG acts as a glucose analogue, and therefore measures the glucose uptake within different organs of the patient. Since tumors have a higher rate of glucose uptake, FDG should be present in higher concentrations than surrounding tissues. However, since all cells require glucose, FDG is not a tumor specific tracer and will also accumulate in healthy organs (Gambhir, 2002).

PET scans are measured by a number known as the standard uptake value (SUV). The SUV at time t is defined as follows.

$$\mathrm{SUV}(t) = \frac{\text{tissue radioactivity concentration}}{\text{injected dose/patient weight}} \tag{1.1.1}$$

The tissue radioactivity concentration is calculated over a certain region of interest (ROI); in our case we will consider this ROI to be a single voxel.

In addition, PET has been explored as a method to guide dose boosting to PET avid regions. Dose boosting accounts for the heterogenity of the tumor in the treatment plan. A uniform dose to a biologically non-uniform tumor is an inefficient use of dose since some voxels will receive more dose than required to obtain tumor control and some voxels may not receive enough dose. In previous studies that utilized PET to guide dose boosting, it was assumed that the values in the image were generally correct. However, PET imaging, like other imaging modalities, is subject to a number of uncertainties. Boellaard (2009) presents a comprehensive list of uncertainties affecting PET imaging. These uncertainties may affect the resulting PET signal making the observed signal not representative of the true underlying biological activity. Thus, any treatment planning paradigm that uses PET imaging to derive heterogeneous prescription doses must also consider the inherent uncertainty in the PET signal.

PET has also been looked at as a method to predict patient outcomes. In particular, high PET-FDG uptake has been associated with worse survival rates for patients (Vansteenkiste et al., 1999; Eschmann

et al., 2006; Downey et al., 2004). It is hoped that the ability to predict patient outcomes based on PET scans can improve the treatment process by allowing clinicians to adapt based on the scan's prediction.

In this thesis, we will investigate the use of PET scans as applied to radiotherapy of lung cancer in two different ways. The first will be to use PET information to guide biologically-based prescription dose. The second will be as a predictive tool to identify patients who will respond to treatment.

1.2 Contributions

Robust biologically-guided intensity-modulated radiation therapy

To the best of our knowledge, there has been no previous work done applying robust optimization to biologically-guided (BG) IMRT treatment planning. In this paper, we develop the first robust optimization approach to PET-based treatment planning. We assume that the PET signal is correlated with hypoxia (lack of oxygen), which in turn produces a heterogeneous spatial distribution of radioresistance, thus requiring a voxel-specific heterogeneous prescription dose. While certain tracers are more suited for measuring hypoxia, our framework is tracer-independent. All that is required is a function to convert the PET image values to a prescription dose. We simply make this assumption as a proof of principle. Noise or uncertainty in the PET signal affects the heterogeneous prescription dose and thus the overall treatment quality; if the true signal is higher than observed, we may be underdosing the tumor and if it is lower than observed, we may miss out on an opportunity for improved sparing. Our approach accounts for uncertainty in the PET signal and extends a previous cardinality-based robust optimization model (Bertsimas and Sim, 2004). We apply our framework to a clinical lung case with PET information. Since we use respiratory-synchronized PET images (4D PET) we do not need to consider the effect of motion blur on the dose-influence matrix.

Lung cancer response classification using pattern recognition

Improved methods to predict response to treatment could lead to therapeutic approaches that are tailored to individual patients. In this work, we leverage gated 4D PET/CT scans to investigate the potential improvements that temporally sequenced data could bring to predict treatment outcome. Past work relied on a population averaged motion kernel to reconstruct unblurred images Vaidya et al. (2011) and only utilized pre-treatment scans. We also investigate the efficacy of different machine learning algorithms on this dataset. To the best of our knowledge, there has been no previous work comparing different machine learning algorithms using FDG-PET to predict patient response.

1.3 Organization

This thesis is structured as follows: In Chapter 2, we present a review of biologically-guided IMRT, approaches to handling uncertainty in IMRT, and PET-guided predictive models found in the literature. In Chapter 3, we present our work in developing a robust optimization framework for biologically-guided IMRT. In Chapter 4, we present our analysis of applying machine learning algorithms to 4D FDG-PET/CT scans. In Chapter 5, we present our concluding remarks and provide directions for future areas of investigation.

Chapter 2

Literature review

In this chapter, we provide an overview and discuss previous research that has been conducted on topics relevant to this thesis. In Section 2.1 we review past work conducted in biologically-guided IMRT. Section 2.2 reviews the literature on uncertainty in radiotherapy (RT) and approaches to combat it. Section 2.3 reviews the literature on uncertainty in biologically guided RT. Finally, in Section 2.4 reviews the use of PET scans in predictive models.

2.1 Biologically-guided IMRT (BGRT)

Biologically-based imaging such as PET can provide insight into factors that may affect the spatial radiosensitivity distribution of the target (Ling et al., 2000). This heterogeneous radiosensitivity may require a non-uniform dose prescription in order to maximize TCP (Brahme and Argren, 1987; Levin-Plotnik and Hamilton, 2004).

How best to utilize biological information to aid treatment planning is still an open question, party due to uncertainties inherent in PET imaging. Optimizing biological objective functions, such as tumor control probability (TCP), rather than physical dose volume objectives, has been investigated. Brahme (1999) conducted early studies on the use of radiobiological objectives in radiation therapy. Kim and Tomé (2006) optimized a biological objective function that maximized TCP while minimizing normal tissue complication probability (NTCP). Kim et al. (2012) proposed a stochastic control method that minimized a nonlinear objective function describing the total number of tumor cells. These approaches directly optimized a nonlinear biological objective function.

An alternative to optimizing the biological objective function directly is to generate a heterogeneous prescription dose distribution. A number of researchers have proposed a tracer-independent linear relationship to transform the PET signal from a biological image to prescription dose (Xing et al., 2002; Alber et al., 2003; Das et al., 2004; Vanderstraeten et al., 2006b,a). These studies were primarily concerned with investigating the feasibility of using PET to introduce a boost dose and did not analyze radiobiological metrics.

Yang and Xing (2005) calculated prescription doses required to maximize TCP given voxel-specific radiobiological parameters under the linear-quadratic (LQ) cell kill model. This approach was then applied to a prostate tumor case using hypothetical distributions of radiobiological parameters. When compared to a uniform dose escalation approach, it was found that a heterogeneous dose distribution

provides an increase in TCP and a decrease in normal tissue complication probability (NTCP).

South et al. (2008) developed a theoretical framework to derive heterogeneous prescription dose distributions based upon functional imaging. A follow up study by South et al. (2009) applied this theoretical framework to PET imaging using FDG as a tracer and investigated the effects of varying resolutions of dose compartments.

Søvik et al. (2006) investigated the use of dose redistribution to target hypoxic compartments of a theoretical tumor model. A four compartment based heterogeneous dose distribution was used. They found that dose redistribution could realize significant increases in TCP compared to an uniform dose. Since TCP is a continuously increasing function of dose, the significance of dose redistribution is that compared to dose escalation, the gains in TCP are due to a more efficient allocation of dose.

2.2 Uncertainties in radiation therapy

Since IMRT treatments are typically given in small doses over a number of fractions and planning is conducted on the pre-treatment CT scan, differences in patient setup between fractions (interfraction) and organ motion during treatment (intrafraction) can introduce errors into the treatment. Current clinical practice is to define a margin around the tumor volume to account for errors in positioning. By extending the tumor volume, this will cause the healthy tissue in the margin area to receive additional dose, potentially impacting the health of the patient.

Much of the previous work in IMRT treatment planning under uncertainty models the effects of patient positioning and organ movement by blurring the dose matrix, resulting in an expected dose rather than a known deterministic dose. Chu et al. (2005) developed a robust optimization model using an ellipsoidal uncertainty set to address interfraction patient shifts and tumor motion. The probabilities for the shift directions were assumed to be known. Their computational results demonstrated that the robust solution achieved comparable target dose homogeneity as the deterministic solution, but was able to realize a reduction in dose to healthy organs. Olafsson and Wright (2006) investigated dose calculation and interfraction patient positioning errors and introduce a novel algorithm, sequential linear programming (SLP), to improve the solution time of their robust second order cone program (SOCP).

Chan et al. (2006) and Bortfeld et al. (2008) considered intrafraction uncertainty due to patient breathing motion. Data from breathing traces was used to construct error bars on the breathing motion. The robust model did not rely on knowledge of underlying probability distribution. The resultant robust treatment plan was able to deliver more dose to the phantom tumor than the nominal plan and less dose to the healthy tissues than the margin plan. Therefore, this robust plan represents a tradeoff between the plan that assumes no uncertainty (nominal) and the one that assumes maximum uncertainty (margin). Robust optimization methods have also been developed for intensity-modulated proton therapy (Unkelbach et al., 2007; Fredriksson et al., 2011; Liu et al., 2012; Chen et al., 2012). Patient anatomy also changes during radiotherapy (Lim et al., 2011). Adaptive methods have been proposed to account for changes in tumor size and position. Adaptive radiotherapy involves modifying the plan based on new data over the course of treatment at certain time intervals.

2.3 Uncertainties in BGRT

PET imaging, like other imaging modalities, is subject to a number of uncertainties. Boellaard (2009) presents a comprehensive list of uncertainties affecting PET imaging, with the three largest being blood glucose level, region of interest, and patient motion, which had maximum effects of $\pm 75\%, \pm 55\%, \pm 60\%$, respectively. Increased blood glucose could result in lowered SUV. The size and shape of the tumor can greatly impact the SUV, with the SUV of small tumors being underestimated more. Patient motion and breathing introduces image artifacts and decreased resolution. Certain uncertainties can be controlled, for example, glucose level can be controlled by having the patient fast before scan acquisition, and patient motion can be controlled by using gated PET scans. However, uncertainties may still affect the resulting PET signal, making the observed signal not representative of the true underlying biological activity. Christian et al. (2009) demonstrated on animal models the limitations of PET imaging and the need to account for discrepancies in the image due to the finite resolution of the scanning technology. Thus, any treatment planning paradigm that uses PET imaging to derive heterogeneous prescription doses must also consider the inherent uncertainty in the PET signal.

A few studies have considered incorporating uncertainty into a biologically-based treatment planning process. Kåver et al. (1999) compared the use of stochastic optimization and margins on the biological parameters to maximize the expected probability of uncomplicated treatment (P_+). Biological margins assume that the patient's healthy organs are more radiosensitive and that the patient's tumor is more radioresistant than population-derived mean radiobiological parameters. The simpler method of biological parameter margins obtained results that were similar to stochastic optimization, but the results relied heavily on the choice of margin magnitude.

Witte et al. (2011) addressed geometric uncertainties when using heterogeneous dose distributions in a probabilistic optimization model, where random and systematic (positional) errors were approximated using isotropic Gaussian kernels. The resulting treatment plan was able to obtain results with 99% of the tumor volume having a maximum underdose of less than 5% at a 90% confidence level.

2.4 PET as a prognostic tool

Studies have shown that FDG has potential in predicting the response of cancer to radiotherapy treatment (Grigsby et al., 2004; Petit et al., 2009). A large body of research has been conducted to identify clinically relevant metrics on which to base a prediction of patient response. Interested readers can refer to Juweid and Cheson (2006) and Ben-Haim and Ell (2009) for a comprehensive literature review of PET-FDG and its role in the evaluation of patient response to cancer treatment.

Goodgame et al. (2008) found that for patients who underwent curative surgical resection for stage 1 non-small cell lung cancer (NSCLC) the pre-operative FDG with a maximum SUV (SUV_{max}) larger than 5.5 predicted relapse and death. Aerts et al. (2009) found that the residual metabolically-active areas in the post-radiotherapy PET-FDG scan corresponded to the areas of high FDG uptake in the original pre-radiotherapy scan. Dooms et al. (2009) showed that SUV_{max} and an SUV_{max} corrected for the partial volume effect was correlated with a higher risk of death under a univariable analysis. Tumor size was also found to be highly correlated with SUV_{max}. A meta-analysis of the medical literature (Berghmans et al., 2008) suggested that the primary tumor SUV has prognostic value.

El Naqa et al. (2009) investigated feature based methods to predict patient response and implemented a logistic regression model to predict loco-regional and regional failure probabilities of cervix and head and neck cancers based on FDG-PET/CT imaging. Vaidya et al. (2011) conducted a follow up to El Naqa's original study, focusing on lung cancer. A population-based motion kernel was implemented to correct for lung motion (Vaidya et al., 2011). However, only pre-treatment PET-CT scans were considered as input into the predictive model. van Stiphout et al. (2011) utilized a linear SVM to predict patient response after chemotherapy for rectal cancer. They also observed that addition of post-chemotherapy PET data resulted in the best predictive model. Tixier et al. (2011) investigated the use of textural features to predict esophageal cancer patient response.

Readers can also refer to Cruz and Wishart (2006) for a comprehensive overview of applications of machine learning in cancer prediction and prognosis. Machine learning and pattern recognition techniques have also been applied extensively to medical fields outside of cancer treatment prediction (Magoulas and Prentza, 2001; Baek et al., 2009; Inza et al., 2010; Suzuki et al., 2012), such as, but not limited to: image reconstruction, diagnosis, and identification of biomarkers.

Chapter 3

Robust biologically-guided intensity-modulated radiation therapy

In this chapter, we develop a robust optimization method that accounts for uncertainty in the observed SUV. The contributions of this chapter are as follows.

1. We develop a robust optimization methodology which considers protecting against uncertainty across rows of constraints. Previous approaches (Bertsimas and Sim, 2004) considered protecting against uncertainty coupled across columns but in single rows independent from each other.

2. We demonstrate computational results on simulated realizations of uncertainty on a clinical lung cancer case which show improvements in objective function value and TCP when our robust formulation is used in BG-IMRT.

The chapter is organized as follows. Section 3.1 describes our radiobiological modelling and robust methodology. Section 3.1.5 describes the optimization parameters and presents some summary statistics about the patient case. Section 3.2 presents our results. Section 3.3 presents the results of our sensitivity analysis on some selected free parameters in the radiobiological models. Section 3.4 presents our discussion of the results. Section 3.5 presents some work using a more biologically motivated source of uncertainty, the partial volume effect.

3.1 Methods and materials

We assume that a voxel-specific standardized uptake value (SUV) obtained from a FDG-PET image is used to generate a heterogeneous dose distribution. We aim to achieve this heterogeneous dose distribution using both non-robust and robust optimization methods. The two methods are then compared through a simulation study.

We acknowledge that FDG is not a direct indicator of hypoxia, but we use it as a proof of concept to illustrate simple patterns of spatially heterogeneous radioresistance. In reality, FDG uptake is believed to be affected by a number of biological processes including radioresistance (Aerts et al., 2009), proliferation

(Vesselle et al., 2000), cell density (Dooms et al., 2009), and hypoxia (Airley et al., 2001; van Baardwijk et al., 2007; Sattler and Mueller-Klieser, 2009). The optimization model we present is tracer-independent and only requires voxel-specific prescription doses. To generate the prescription dose for each voxel, we start with an observed SUV value, convert it to partial oxygen pressure, generate radiosensitivity parameters, and finally compute the voxel-specific prescribed dose needed to achieve a certain TCP (see Figure 3.1).

3.1.1 Radiobiological modeling

We calculate the tumor control probability of voxel i, TCP_i, according to linear quadratic (LQ) model of cell kill (Fowler, 1989; Webb and Nahum, 1999; Yang and Xing, 2005; South et al., 2009). TCP_i is given by equation (3.1.1) :

$$\mathrm{TCP}_i = \exp\left[\rho_i V_i \exp\left(-\alpha_i d_i - \beta_i \frac{d_i^2}{r} + \frac{\log 2}{T_p}(T - T_k)\right)\right]. \tag{3.1.1}$$

Parameters α_i and β_i are the linear and quadratic coefficients of the LQ model for voxel i, respectively. Parameter T_p is potential cell doubling time in days, T is the overall treatment time, T_k is the number of days until repopulation begins, ρ_i is the number of clonogens per cm^3 in voxel i, V_i is the volume of voxel i, r is the number of fractions, and d_i is the total dose to a voxel i. The TCP for the CTV is calculated as the product of TCP_i for each voxel i in the CTV.

$$\mathrm{TCP} = \prod_{i \in \mathrm{CTV}} \mathrm{TCP}_i \tag{3.1.2}$$

3.1.2 Conversion from SUV to voxel-specific prescription dose

While the true function to convert FDG SUV to a prescription dose is unknown, we will assume a particular functional form for the purposes of model evaluation. The optimization is independent of the prescription dose function; it only assumes that a prescription dose can be calculated given the SUV. Dose prescriptions will be derived to maximize the TCP under the LQ model.

Following South et al. (2009), we assume that the PET signal is negatively correlated to oxygenation. First, we convert SUV θ_i, to oxygenation $p(\theta_i)$, according to the following equation:

$$p(\theta_i) = \min\left\{\max\left\{\phi_1 \exp\left(\frac{\phi_2}{\theta_i}\right), p_{\min}\right\}, p_{\max}\right\} \tag{3.1.3}$$

Parameters $p_{\min}$ and $p_{\max}$ were chosen to be 1 and 100, respectively. The function $p(\theta_i)$ is shown in Figure 3.2a. We chose parameters ϕ_1 and ϕ_2 to fit the overall relationship between SUV and dose so that at a SUV of 2.5 the dose would be approximately 65 Gy and at a SUV of 20, the dose would be approximately 100 Gy (see Figure 3.2d). The values chosen for ϕ_1, ϕ_2 were 1 and 11.5, respectively. The value of 2.5 was used because an SUV of 2.5 is a commonly used threshold for tumor edge deliniation (Hellwig et al., 2007).

For each voxel i, we convert partial oxygen pressure into the factors, A_i and B_i, which modify the intrinsic radiosensitivity parameters, α_i and β_i, respectively (Wouters and Brown, 1997), according to

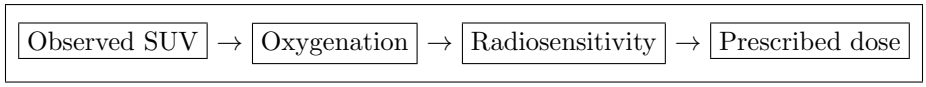
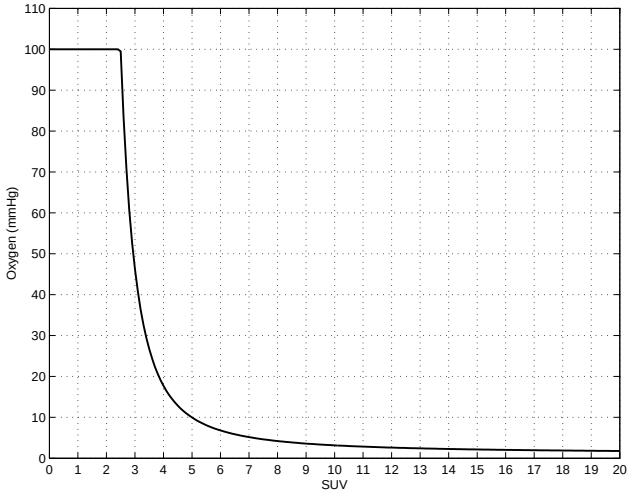
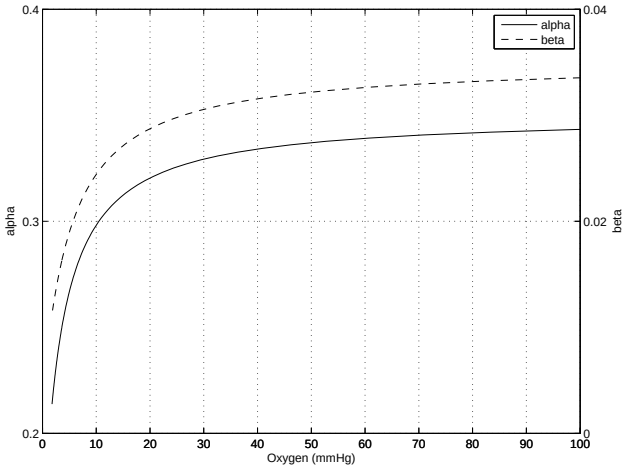


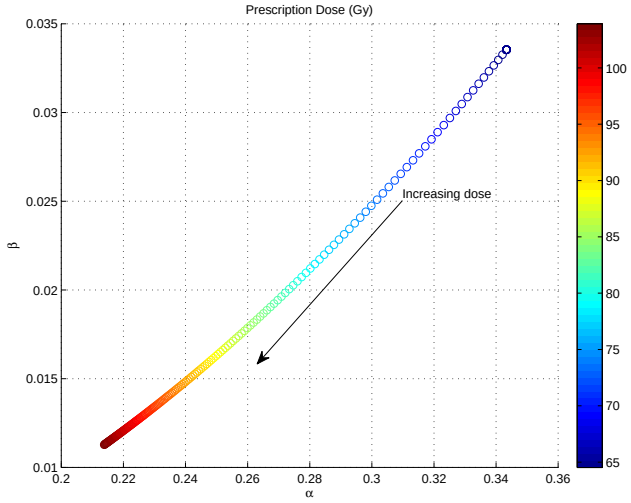
Figure 3.1: Dose conversion process



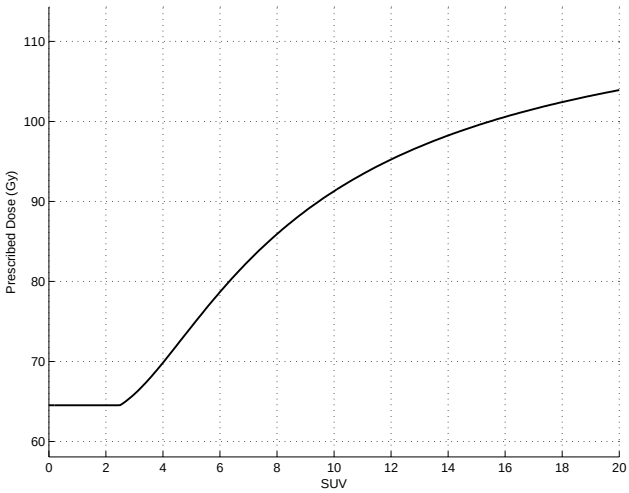
(a) SUV to partial oxygen pressure



(b) Partial oxygen pressure to radiosensitivity parameters



(c) Radiosensitivity parameters to prescribed dose



(d) SUV to prescribed dose

Figure 3.2: Conversion functions

the following equations:

$$A_i = \left(\frac{1}{\mathrm{OER}_{\alpha_{\max}}}\right)\left[\frac{(p_i \mathrm{OER}_{\alpha_{max}} + K_m)}{p_i + K_m}\right] \tag{3.1.4}$$

$$B_i = \left(\frac{1}{\mathrm{OER}_{\beta_{\max}}}\right)^2\left[\frac{(p_i \mathrm{OER}_{\beta_{max}} + K_m)}{p_i + K_m}\right]^2 \tag{3.1.5}$$

Parameters $\mathrm{OER}_{\alpha_{\max}}$ and $\mathrm{OER}_{\beta_{\max}}$ are the maximum oxygen enhancement ratios (OER) for α and β, respectively (Toma-Dasu et al., 2009). Parameter K_m is the partial oxygen pressure at which half-maximum sensitization is reached and p_i is the partial oxygen pressure for voxel i. To derive voxel-specific radiosensitivity parameters α_i and β_i, the intrinsic parameters, α and β, are then multiplied by their respective A_i and B_i: $\alpha_i = A_i\alpha$ and $\beta_i = B_i\beta$. Figure 3.2b shows the relationship between oxygenation and the radiosensitivity parameters.

The modified radiosensitivity parameters determine a voxel-specific prescription dose that results in a specified TCP, according to the method described by Yang and Xing (2005). Figure 3.2c shows the conversion from the radiosensitivity parameters to a voxel-specific prescription dose that results in a voxel TCP of 0.99999. Figure 3.2d summarizes all the steps into a single plot, showing the relationship between the input SUV and output prescription dose.

For this study, the intrinsic radiobiological parameters for a lung tumor were set as follows (Mehta et al., 2001; El Sharouni et al., 2005; South et al., 2009): $\rho = 10^7$ clonogens/cm^3, $\alpha = 0.35$ Gy^{-1}, $\beta = 0.035$, $\frac{\alpha}{\beta} = 10$, $T_p = 3$ days (equivalent to a cell proliferation rate of 0.1386 days^{-1}), $\mathrm{OER}_{\alpha_{\max}} = 2.5$, $\mathrm{OER}_{\beta_{\max}} = 3$, $K_m = 3.28$, $T_k = 28$ days. Partial oxygen pressure for a voxel was assumed to be bounded within the range $p \in [1, 100]$ mm Hg (South et al., 2009). The treatment plan parameters were set as follows: $T = 44$ days, $r = 32$ fractions.

3.1.3 Model of SUV uncertainty

We assume that the true SUV is not directly observed due to noise in the image (Boellaard, 2009). We define θ_i to be the nominal (i.e., observed) SUV in voxel i of the PET image. The true (i.e., unobserved) SUV for voxel i, $\tilde{\theta}_i$, resides in an interval $\mathcal{U}_i = [\theta_i - \hat{\theta}_i, \theta_i + \hat{\theta}_i]$, where $\hat{\theta}_i$ is the maximum absolute deviation from the nominal SUV for voxel i. In other words, the true SUV equals θ_i plus an uncertain noise term.

The values θ_i and $\hat{\theta}_i$ will be used to determine a prescription dose for treatment planning. Since the true SUV value is unknown, the dose that should have been prescribed may be different from the nominal prescription dose, which is derived from the observed SUV. Thus, if we design a treatment using the nominal heterogeneous prescription dose distribution, we may underdose or overdose certain voxels with respect to the dose that should have been prescribed.

3.1.4 Mathematical formulation

We develop a robust optimization framework for treatment planning in the presence of PET signal uncertainty based on the "budget of protection" model of Bertsimas and Sim (2004). Our goal is to demonstrate a proof of concept for the value of robustness in the presence of PET signal uncertainty. As such, we use a simple penalty-based linear model that approximates a treatment planning formulation, but omits many of the more sophisticated features of clinical formulations for the sake of simplicity.

Let i, j, and k index voxels, beamlets, and structures, respectively. Let $\mathcal{I}_k$ be the set of all voxels in structure k, $\mathcal{O}$ be the index set of all organs-at-risk (OARs), $\mathcal{I}_\mathcal{T}$ be the set of all tumor voxels, $\mathcal{I}_\mathcal{O}$ be the set of all OAR voxels, and $\mathcal{J}$ be the set of all beamlets. Let x_i, y_i, and z measure the overdose to voxel i, underdose to voxel i, and maximum underdose to the CTV, respectively. Let w_j be the intensity of beamlet j. Let D_{ij} be the influence matrix describing the dose from unit intensity of beamlet j to voxel i. Let $f(\theta_i)$ be the prescribed dose to voxel i, given SUV θ_i. Let U_i and L_i be the upper and lower bounds on the dose to voxel i, respectively. Let μ_k be an upper bound on the mean dose to OAR k. We will refer to Formulation (3.1.6) as the nominal (no uncertainty) model.

$$\underset{w,x,y,z}{\text{minimize}} \quad \lambda^+ \sum_{i\in\mathcal{I}_\mathcal{T}} x_i + \lambda^- \sum_{i\in\mathcal{I}_\mathcal{T}} y_i + \lambda z + \lambda^\mathcal{O} \sum_{i\in\mathcal{I}_\mathcal{O}} \sum_{j\in\mathcal{J}} D_{ij} w_j \tag{3.1.6a}$$

$$\text{subject to} \quad -y_i \le \sum_{j\in\mathcal{J}} D_{ij} w_j - f(\theta_i) \le x_i, \quad \forall i \in \mathcal{I}_\mathcal{T}, \tag{3.1.6b}$$

$$z \ge y_i, \quad \forall i \in \mathcal{I}_\mathcal{T}, \tag{3.1.6c}$$

$$L_i \le \sum_{j\in\mathcal{J}} D_{ij} w_j \le U_i, \quad i \in \mathcal{I}_\mathcal{O}, i \in \mathcal{I}_\mathcal{T}, \tag{3.1.6d}$$

$$\frac{1}{|\mathcal{I}_k|} \sum_{i\in\mathcal{I}_k} \sum_{j\in\mathcal{J}} D_{ij} \le \mu_k, \quad \forall k \in \mathcal{O}, \tag{3.1.6e}$$

$$x_i, y_i, w_j \ge 0, \quad \forall i \in \mathcal{I}_\mathcal{T}, j \in \mathcal{J}. \tag{3.1.6f}$$

The objective (3.1.10a) minimizes a weighted combination of the total absolute deviation from the voxel-specific prescribed dose for all target voxels and the total dose to the OARs. Parameters $\lambda^+, \lambda^-, \lambda, \lambda^\mathcal{O}$ are the weights for penalizing overdose, underdose, maximum underdose, and OAR dose, respectively. Constraint (3.1.6b) models the overdose and underdose computations. Constraint (3.1.6c) models the maximum underdose computation. Constraint (3.1.6d) bounds the upper and lower doses to every OAR voxel. Constraint (3.1.6e) limits the mean dose for each OAR. Since Formulation (3.1.6) assumes no uncertainty in the underlying radiobiological parameters, a treatment that is able to deliver at least the prescribed dose to all CTV voxels will generate a voxel TCP of 0.99999 and an overall CTV TCP of at least $0.99999^{|\mathcal{I}_{\text{CTV}}|}$, where $|\mathcal{I}_{\text{CTV}}|$ is the cardinality of the set of CTV voxels.

The robust formulation accounts for SUV uncertainty and builds on Formulation (3.1.6). First, let $f(\tilde{\theta}_i)$ be the dose prescription for voxel i associated with the uncertain SUV $\tilde{\theta}_i$, which we use in place of $f(\theta_i)$ in constraint (3.1.6b). Second, to facilitate the comparison of our robust model with the nominal model, we require the robust model to deliver a mean dose to the tumor that is the same as what is delivered in the nominal model. Thus, after solving the nominal model we calculate the mean tumor dose, which we define as Ω, and include a constraint which holds the mean CTV dose delivered by the robust solution equal to Ω. These two changes to the nominal model are implemented by eliminating constraint (3.1.6b) in Formulation (3.1.6) and adding the following two constraints:

$$-y_i \le \sum_{j\in\mathcal{J}} D_{ij} w_j - f(\tilde{\theta}_i) \le x_i, \quad \forall \tilde{\theta}_i \in \mathcal{U}_i, i \in \mathcal{I}_\mathcal{T}, \tag{3.1.7a}$$

$$\frac{1}{|\mathcal{I}_\mathcal{T}|} \sum_{i\in\mathcal{I}_\mathcal{T}} \sum_{j\in\mathcal{J}} D_{ij} w_j = \Omega. \tag{3.1.7b}$$

Since $\mathcal{U}_i$ is an interval, a formulation including the constraint (3.1.7a) is not directly solvable as a

linear program. However, we can easily reformulate it into a tractable linear program. Our model assumes that each voxel's SUV may deviate from its nominal value; however, the total number of voxels that will deviate (and conspire to produce a worst-case effect) is bounded. We introduce a parameter $\gamma_i \in [0,1]$ for each constraint involving $\mathcal{U}_i$ to model to the extent of the SUV change of voxel i. For example, $\gamma_i = 0$ models the belief that voxel i will remain at its nominal SUV θ_i and therefore only require the corresponding nominal prescription dose, while $\gamma_i = 1$ models the belief that voxel i will realize its worst-case (largest) SUV $\theta_i + \hat{\theta}_i$ and therefore require a higher dose to achieve the same TCP. Formulation (3.1.8) shows the linear robust formulation.

$$
\begin{aligned}
\underset{w,x,y,z}{\text{minimize}} \quad & \lambda^+ \sum_{i\in\mathcal{I}_\mathcal{T}} x_i + \lambda^- \sum_{i\in\mathcal{I}_\mathcal{T}} y_i + \lambda z + \lambda^\mathcal{O} \sum_{i\in\mathcal{I}_\mathcal{O}} \sum_{j\in\mathcal{J}} D_{ij} w_j \\
\text{subject to} \quad & \sum_{j\in\mathcal{J}} D_{ij} w_j - f(\theta_i) + \gamma_i \left[f(\theta_i + \hat{\theta}_i) - f(\theta_i) \right] \leq x_i, \quad \forall i \in \mathcal{I}_\mathcal{T}, \\
& -\sum_{j\in\mathcal{J}} D_{ij} w_j + f(\theta_i) + \gamma_i \left[f(\theta_i + \hat{\theta}_i) - f(\theta_i) \right] \leq y_i, \quad \forall i \in \mathcal{I}_\mathcal{T}, \\
& z \geq y_i, \quad \forall i \in \mathcal{I}_\mathcal{T}, \\
& L_i \leq \sum_{j\in\mathcal{J}} D_{ij} w_j \leq U_i, \quad i \in \mathcal{I}_\mathcal{O}, i \in \mathcal{I}_\mathcal{T}, \\
& \frac{1}{|\mathcal{I}_k|} \sum_{i\in\mathcal{I}_k} \sum_{j\in\mathcal{J}} D_{ij} \leq \mu_k, \quad \forall k \in \mathcal{O}, \\
& \frac{1}{|\mathcal{I}_\mathcal{T}|} \sum_{i\in\mathcal{I}_\mathcal{T}} \sum_{j\in\mathcal{J}} D_{ij} w_j = \Omega, \\
& x_i, y_i, w_j, \geq 0, \quad \forall i \in \mathcal{I}_\mathcal{T}, j \in \mathcal{J}.
\end{aligned}
\tag{3.1.8}
$$

If we allow γ_i to be decision variables that can be optimized for simultaneously with the beamlet intensities w_j, the optimization engine will chose values for γ_i in an optimistic (i.e., not worst-case) manner. Instead, to ensure that Formulation (3.1.8) is protecting against a worst-case realization of the SUV, we formulate an auxiliary optimization problem (3.1.9) that identifies the worst-case combination of voxels to change SUV and chooses the γ_i values accordingly within an overall budget of Γ.

$$
\begin{aligned}
\underset{\gamma}{\text{maximize}} \quad & \sum_{i\in\mathcal{I}_\mathcal{T}} \gamma_i \left[f(\theta_i + \hat{\theta}_i) - f(\theta_i) \right] \\
\text{subject to} \quad & \sum_{i\in\mathcal{I}_\mathcal{T}} \gamma_i \leq \Gamma, \\
& 0 \leq \gamma_i \leq 1, \quad \forall i \in \mathcal{I}_\mathcal{T}.
\end{aligned}
\tag{3.1.9}
$$

The value Γ can be interpreted as the maximum number of voxels that we expect to change their SUV to their worst-case values. Therefore Γ need not be larger than $\Gamma_{\max} := |\mathcal{I}_\mathcal{T}|$ (the number of tumor voxels). Choosing $\Gamma = 0$ forces all $\gamma_i = 0$, and the robust model (3.1.8) reduces to the nominal model (3.1.6). Choosing $\Gamma = |\mathcal{I}_\mathcal{T}|$ results in the most conservative model, where we assume all tumor voxels can change their SUV to their worst-case values. Given a particular Γ, the auxiliary problem (3.1.9) is solved first to find the optimal γ_i^* values, which are then used as input to formulation (3.1.8). Note that formulation (3.1.9) is an instance of the continuous knapsack problem, for which it is known that the greedy solution is optimal. That is, if the values of $\left[f(\theta_i + \hat{\theta}_i) - f(\theta_i) \right]$ are ranked in descending order, the optimal solution is $\gamma_i^* = 1$ for $i = 1, \ldots \lfloor \Gamma \rfloor$ and $\gamma_{\lceil \Gamma \rceil}^* = \Gamma - \lfloor \Gamma \rfloor$ (in the case Γ is not integer). Hence,

the effort required to solve formulation (3.1.9) is trivial. Overall, the parameter Γ allows a treatment planner to adjust his or her level of conservatism when designing the robust treatment.

Differences between our row-based approach and a column-based approach

We mentioned earlier in the introduction to this chapter that our robust approach differs from the previous robust approach developed by Bertsimas and Sim (2004). Their robust model is formulated as follows. Let J_i be the set of coefficients a_{ij} in row i and column j which are subject to uncertainty. This model also assumes that the true value of a_{ij} resides in the interval $[a_{ij} - \hat{a}_{ij}, a_{ij} + \hat{a}_{ij}]$.

$$\underset{\mathbf{x},\mathbf{y}}{\text{maximize}} \quad \mathbf{c}'\mathbf{x} \tag{3.1.10a}$$

$$\text{subject to} \quad \sum_j a_{ij}x_j + \max_{\{S_i \bigcup \{t_i\} | S_i \subseteq J_i, |S_i| = \lfloor \Gamma_i \rfloor, t_i \in J_i \setminus S_i\}} \left\{ \sum_{j \in S_i} \hat{a}_{ij} y_j + (\Gamma_i - \lfloor \Gamma_i \rfloor)\, \hat{a}_{it_i} y_t \right\} \leq b_i, \quad \forall i, \tag{3.1.10b}$$

$$-y_j \leq x_j \leq y_j, \quad \forall j, \tag{3.1.10c}$$

$$\mathbf{l} \leq \mathbf{x} \leq \mathbf{u} \tag{3.1.10d}$$

$$\mathbf{y} \geq 0. \tag{3.1.10e}$$

Constraint 3.1.10b is equivalent to the following linear program.

$$\begin{aligned} \underset{\mathbf{z}_i}{\text{maximize}} \quad & \sum_{j \in J_i} \hat{a}_{ij} |x_j^*| z_{ij} \\ \text{subject to} \quad & \sum_{j \in J_i} z_{ij} \leq \Gamma_i, \\ & 0 \leq z_{ij} \leq 1, \quad \forall j \in J_i. \end{aligned} \tag{3.1.11}$$

Model 3.1.11 is then embedded into Model 3.1.10 by taking the dual of Model 3.1.11. The proofs of equivalence and details of conversion are omitted from this thesis; interested readers can refer to the original paper (Bertsimas and Sim, 2004).

In constrast to our model, there exists a parameter Γ_i for all rows i. Γ_i is a budget of uncertainty for row i and it protects against the coefficients a_{ij} from a change in their nominal value. The selection of optimal $\hat{a}_{ij}$ to protect against is determined through the solution of Model 3.1.11. Our model only has a single parameter Γ, which couples the protection across the rows of the model.

3.1.5 Patient data and optimization parameters

The data used in this study was obtained from The Princess Margaret Cancer Center. Combined 4D FDG-PET/CT images and tumor contours were exported from Philips Pinnacle in DICOM format. The data was then imported into the Computational Environment for Radiotherapy Research (CERR), in order to generate the dose influence matrices necessary for optimization (Deasy et al., 2003). The SUV of each voxel from the PET image was converted into a voxel-specific prescription dose as described in Section 3.1.2.

We penalized underdosing more heavily than overdosing on the target prescription dose since underdosing (especially maximum underdose) has a larger effect on TCP. Penalty parameters were set as

	SUV		Prescription dose	
	PTV\CTV	CTV	PTV \CTV	CTV
Mean	0.97	3.61	64.73	71.23
Max	8.97	18.62	88.73	102.91
Min	0.03	0.12	64.53	64.53
Std	0.81	3.68	1.52	10.36
Median	0.76	1.93	64.53	64.53

Table 3.2: Summary statistics of SUV and prescription dose distribution

follows: $\lambda^+ = 0, \lambda^- = 1000, \lambda = 100,000, \lambda^{\mathcal{O}} = 1$. Mean and maximum dose constraints were obtained from the QUANTEC series of papers (Marks et al., 2010) and clinical protocols at Princess Margaret Cancer Center (see Table 3.3). For organs without a clinical or QUANTEC specified maximum limit, we set the maximum to 100. For the target, we set an upper limit of 1.4 times the prescription dose at each voxel. Table 3.1 outlines the parameters for mean and max constraints in the nominal and robust models.

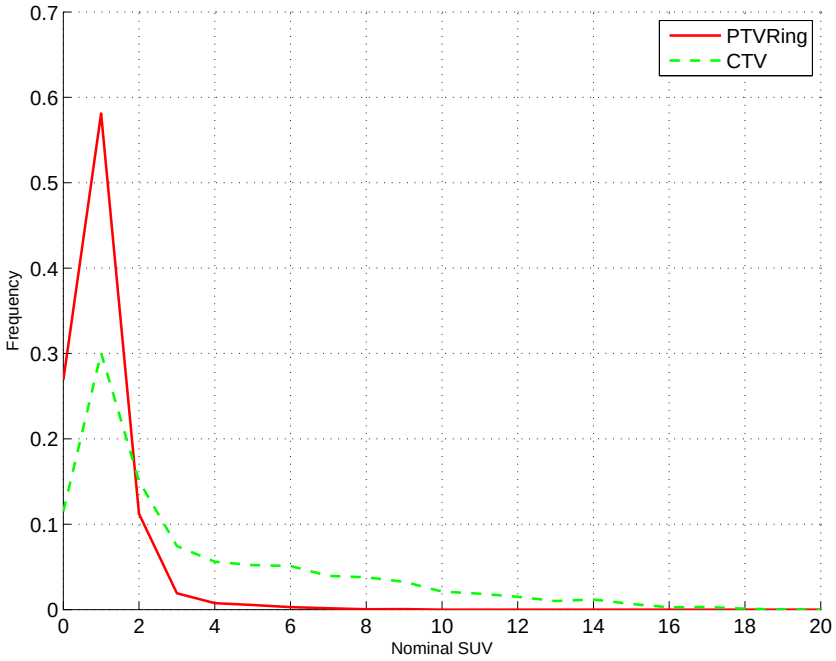
	Lung	Esophagus	Spinal Canal	Heart	Target
Mean dose	18	34	-	26	-
Max dose	100	65	50	100	$1.4f(\theta_i)$

Table 3.1: Model mean and maximum dose limits (Gy)

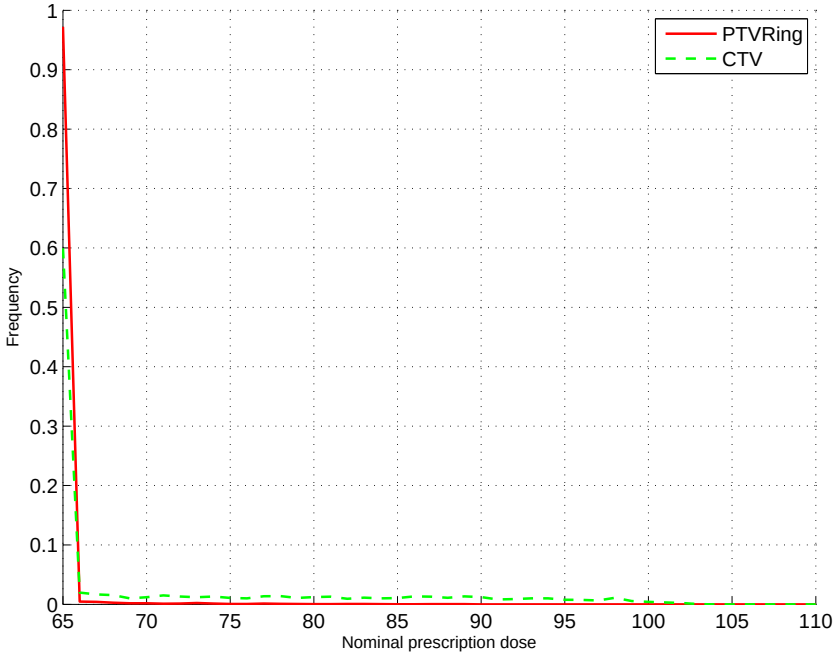
For the robust formulation, we set $\Omega = 83.25$ Gy, which is the mean tumor dose of the solution to the nominal formulation. We employed a coplanar equispaced 7 beam configuration for the two optimization models. The models were solved using CPLEX 12 on a remote computing cluster. The node used was composed of a 2.27 Ghz Intel Xeon 10 core processor and 252 Gb of RAM. All models solved within 1 minute.

The original CT scan had a resolution of 0.0977 cm $\times$ 0.0977 cm $\times$ 0.2 cm. The original PET scan had a resolution of 0.3906 cm $\times$ 0.3906 cm $\times$ 0.3270 cm. The SUV of the PET voxels were mapped to their closest corresponding voxel in the CT image. Then the CT scan was downsampled to obtain a final planning resolution of 0.3906 cm $\times$ 0.3906 cm $\times$ and 0.2 cm. This resulted in 7,372 voxels in the CTV and 207,938 voxels in the OARs. Since the prescription dose is generated to obtain a voxel TCP of 0.99999, a plan that is able to deliver at least the prescribed dose to all CTV voxels will obtain an overall CTV TCP of 0.93 (0.99999^{7372}).

Figure 3.3a shows the distribution of the SUV in different regions of interest within the target. It can be observed that the portion of the PTV that is not part of the CTV tends to have most of its SUV clustered around values between 0-2. While the CTV itself has a larger range of values and is more spread out. Figure 3.3b shows the distribution of prescription dose in different regions of interest in the target. It can be observed that the non-CTV PTV ring has nearly all of its prescription dose values at 65 Gy, while the CTV itself has some small hotspots of high prescription dose and a large portion of voxels at the standard 65 Gy. Table 3.2 shows the summary statistics for both SUV and prescription dose values. Figure 3.4 shows the spatial distribution of the PET scan's SUV uptake at a specified z-axis coordinate.

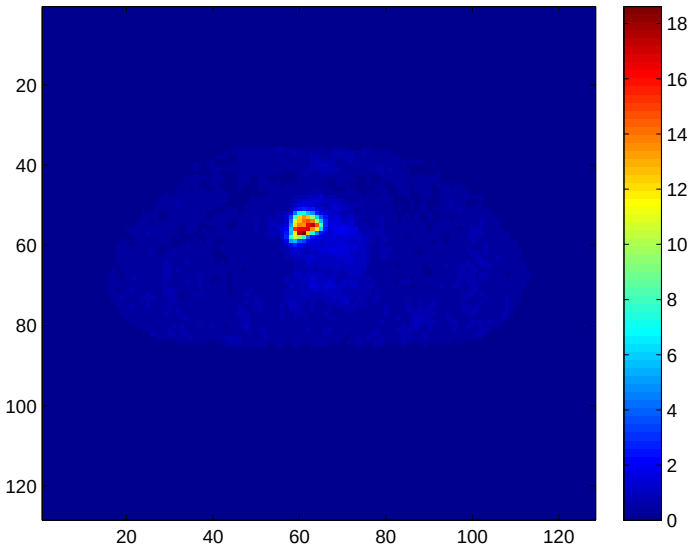


(a) Distribution of nominal SUV

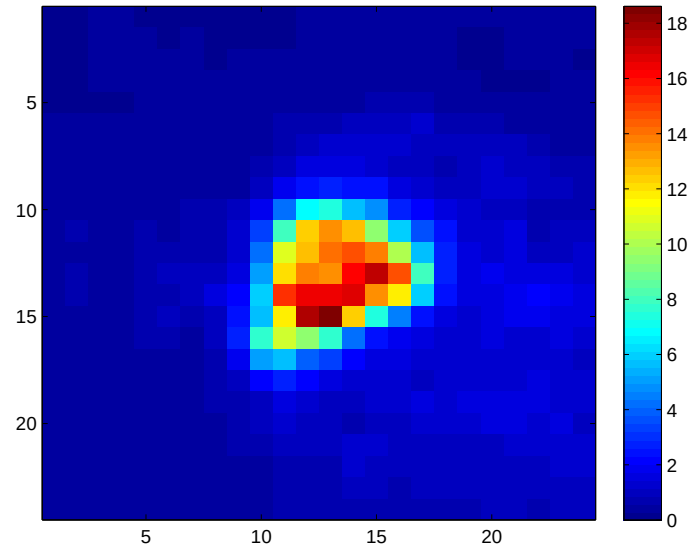


(b) Distribution of nominal prescription dose

Figure 3.3: Distribution of nominal SUV and prescription dose



(a) 2D slice of the PET scan.



(b) Magnified 2D slice of the PET scan.

Figure 3.4: 2D slice showing the spatial distribution of SUV uptake. The colors represent the SUV at that voxel.

3.1.6 Simulating SUV to evaluate optimization results

Once we determine optimal beamlet intensities by solving the nominal and robust formulations, we simulate the realization of different SUV and evaluate the dosimetric consequences. We conduct two types of computational analyses which are described as follows.

Random

The first computational analysis considers an abstract random simulation of SUV realization. This could be thought of as random noise in the image or as the confounding of many different sources of uncertainty some of which bias the SUV positively and some which bias it negatively, leading to a "random" realization of the observed image value. This analysis is further divided into two different methods of generating SUV realizations.

- Uniform uncertainty scenario: The SUV for voxel i is drawn from a uniform distribution over the interval $[\theta_i - \hat{\theta}_i, \theta_i + \hat{\theta}_i]$.
- Worst case uncertainty scenario: We simply choose the worst-case SUV (the maximum value in the interval).

Each voxel is considered independently of each other. The range of SUV for voxel i was generated by setting $\hat{\theta}_i$ to $0.6\theta_i$. A random set of voxels realize a change from their nominal values in each simulation. We simulated 2000 realizations for each scenario. Section 3.2 describes the results for this simulation.

Partial volume effect (PVE)

The second is a more biologically realistic simulation motivated by the partial volume effect (PVE). This analysis considers PVE in isolation from any other source of uncertainty. Section 3.5 describes the nature of the effect in further detail and also the results of the simulation.

3.2 Results

First, we consider the impact of uncertainty and protection level Γ on the objective function values (i.e., a weighted measure of deviation from prescription dose) and TCP. Recall that $\Gamma = 0$ is equivalent to the nominal plan, while positive Γ values correspond to robust solutions. Figures 3.5 and 3.6 depict histograms of the objective function values and TCP values, respectively, where the relative frequencies are generated from simulating random sets of voxels to experience a change from their nominal SUV. The change in the observed SUV was considered under two scenarios: uniform and worst-case uncertainty. The uniform scenario is when the SUV is allowed to increase or decrease within the bounded range according to a uniform distribution. The worse-case scenario is when the SUV reaches its maximum value within the bounded range. Each voxel is considered independently from each other. TCP was calculated only on the CTV since the use of TCP is not applicable to the PTV, which is a geometric concept used to deal with setup errors (Li et al., 2012).

Figure 3.5 shows that under both the uniform and worst-case uncertainty scenarios, the nominal plan generates worse objective function values and has more variability in these values, compared to the robust plans. A higher objective function value means the delivered dose distribution deviates more from the desired dose distribution. Increased spread in the histogram means that the performance of the

nominal plan is much more variable in the presence of uncertainty. Note for $\Gamma > 0.4\Gamma_{\max}$ the resulting histogram is essentially the same as the one for $\Gamma = 0.4\Gamma_{\max}$. This suggests that only some protection is required to get most of the gains of robustness.

Figure 3.6 shows similar results as Figure 3.5 but from the perspective of TCP. The robust plans generally have higher TCP values and less variability in TCP under the worst-case uncertainty scenarios. Similar to the previous case, negligible gains in TCP are realized when Γ is increased above $0.4\Gamma_{\max}$. Under the uniform uncertainty scenario, the differences between the performance of the robust and nominal plans is less, but the robust plan still provides improvements over the nominal plan.

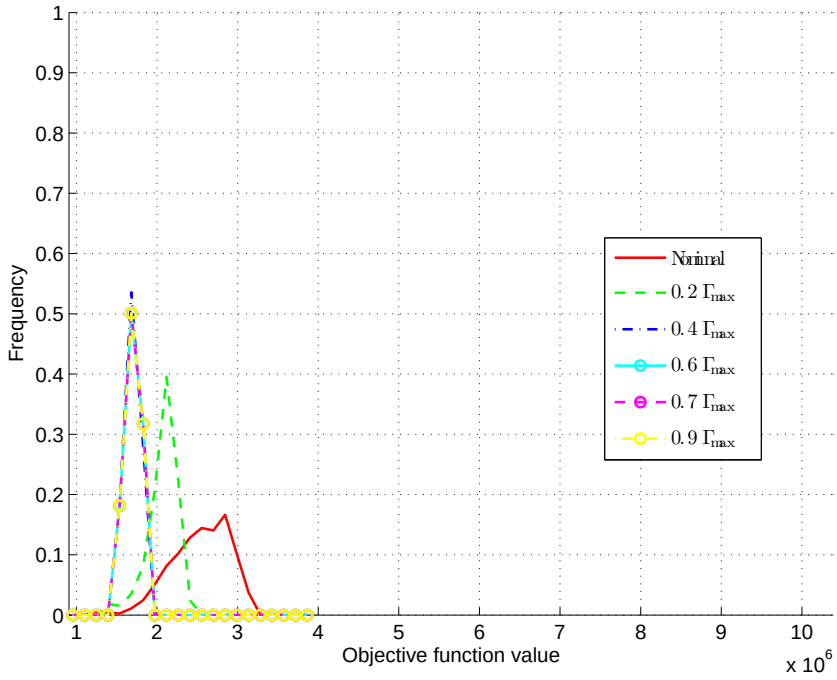
Table 3.3 provides more granular results for the performance of the nominal and robust plans with respect to OAR dose. V_x is the fractional volume of the region of interest (ROI) that exceeds a dose of x Gy. Max cc $\geq x$ Gy is the maximum volume in cubic centimeters of the ROI that is allowed to exceed x Gy. The first row of the table shows the DVH criteria currently used at Princess Margaret Cancer Center. The remaining rows show the DVH values corresponding to the robust plans for varying levels of Γ. For this instance, we were able to design treatment plans that were within preferred clinical limits using both the nominal and robust models.

	Lung				Esophagus		Spinal Canal	Heart	
	Mean (Gy)	V_5	V_{10}	V_{20}	V_{40}	Max (Gy)	Max (Gy)	V_{40}	cc ≥ 74
Clinical	18	0.5	0.4	0.3	0.6	65	50	0.6	2
$0\Gamma_{\max}$	7.62	0.21	0.16	0.11	0.31	65.00	50.00	0.01	0.00
$0.1\Gamma_{\max}$	7.78	0.21	0.16	0.12	0.32	65.00	50.00	0.01	0.00
$0.2\Gamma_{\max}$	8.32	0.25	0.18	0.12	0.32	65.00	50.00	0.01	0.00
$0.3\Gamma_{\max}$	8.55	0.26	0.19	0.13	0.37	65.00	50.00	0.01	0.00
$0.4\Gamma_{\max}$	8.84	0.28	0.21	0.13	0.33	65.00	50.00	0.01	0.00
$0.5\Gamma_{\max}$	8.90	0.28	0.21	0.14	0.34	65.00	50.00	0.01	0.00
$0.6\Gamma_{\max}$	8.90	0.28	0.21	0.14	0.34	65.00	50.00	0.01	0.00
$0.7\Gamma_{\max}$	8.90	0.28	0.21	0.14	0.34	65.00	50.00	0.01	0.00
$0.8\Gamma_{\max}$	8.90	0.28	0.21	0.14	0.34	65.00	50.00	0.01	0.00
$0.9\Gamma_{\max}$	8.90	0.28	0.21	0.14	0.34	65.00	50.00	0.01	0.00
$\Gamma_{\max}$	8.90	0.28	0.21	0.14	0.34	65.00	50.00	0.01	0.00

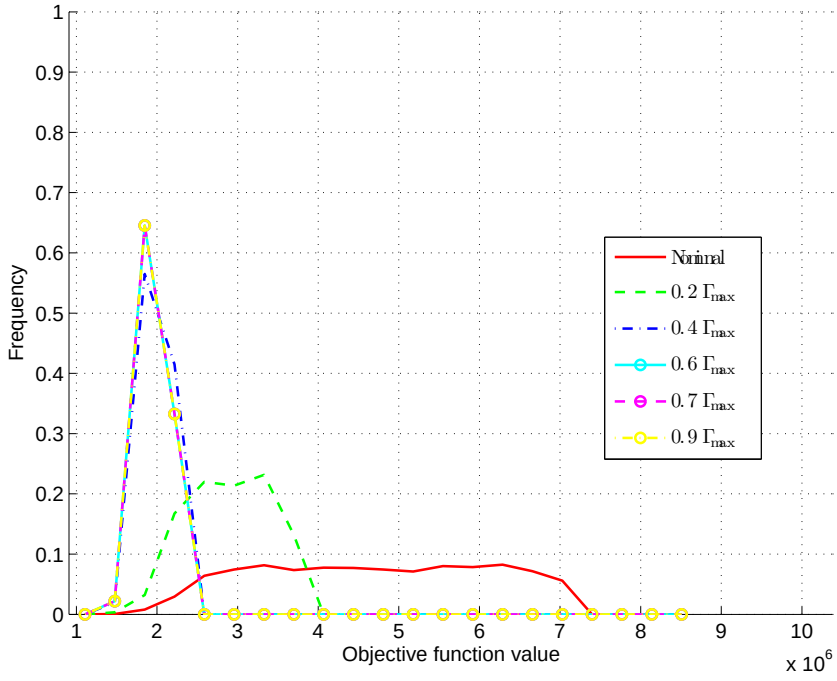
Table 3.3: Computed DVH metrics over varying Γ values

Figure 3.7 compares the DVH of the nominal plan to multiple robust plans (from $\Gamma = 0.1\Gamma_{\max}$ to $\Gamma = \Gamma_{\max}$). We can observe that the robust model delivers an increased dose to the lung and esophagus. The heart DVH experiences little change between the nominal and robust plans. The spinal canal DVH shows a decrease in higher dose volume but an increase in lower dose volume.

In general, there appears to be a trade-off between TCP and mean dose to the lung (Figure 3.8). The x-axis is the mean dose to all OARs. The dose to the lung only changes with the value of Γ since the optimization problem is solved with new parameters and not with each simulated realization of SUV. The y-axis is the mean TCP value over all simulated realizations. As previously observed, there is an increase in minimum TCP and decrease in variance as we increase the value of Γ. However, we also increase the overall dose to the lung. Therefore, a robust plan does not come for free; we incur the trade-off of increasing dose to the lung when we increase the conservativeness of the plan.

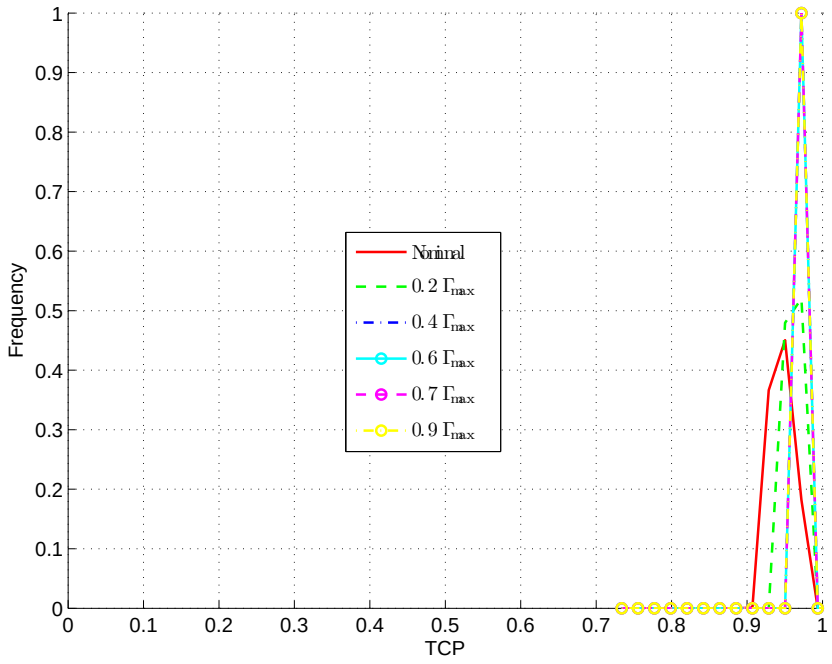


(a) Uniform uncertainty

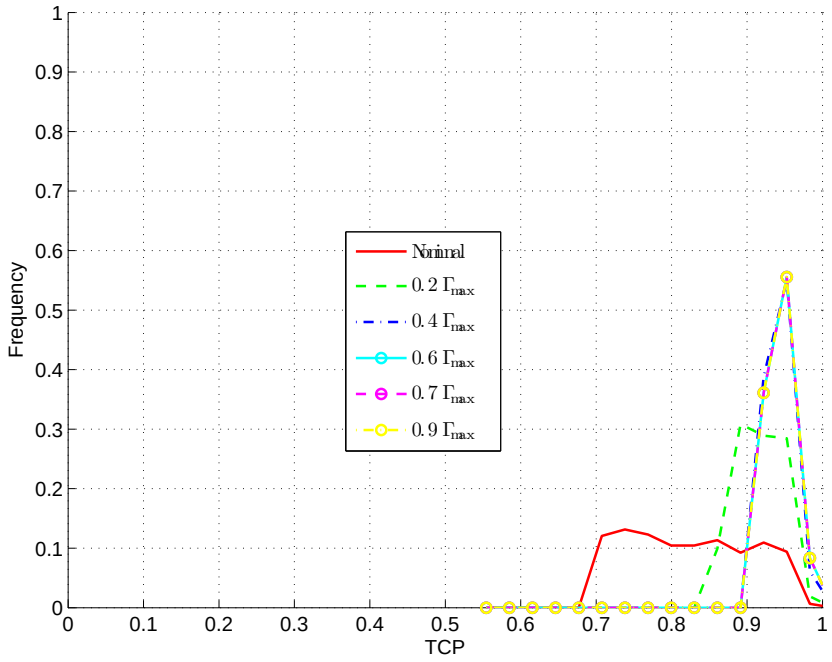


(b) Worst case uncertainty

Figure 3.5: Comparison of realized objective function values between nominal and robust plans



(a) Uniform uncertainty



(b) Worst case uncertainty

Figure 3.6: Comparison of realized TCP values between nominal and robust plans

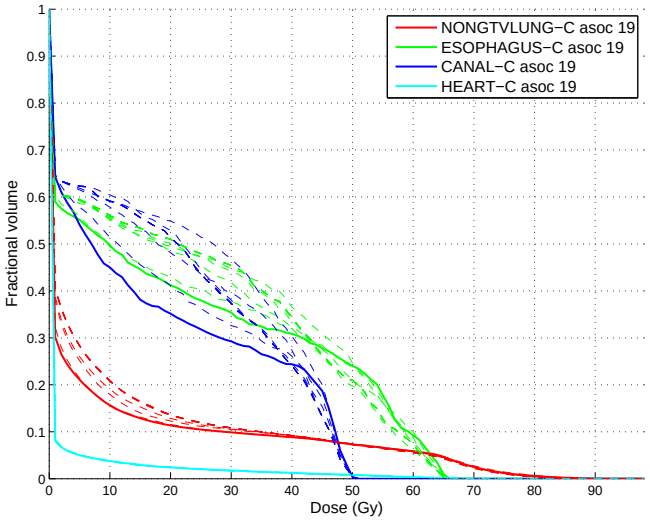


Figure 3.7: Comparison of nominal and robust DVH. The dashed lines represent the DVH of the robust solutions.

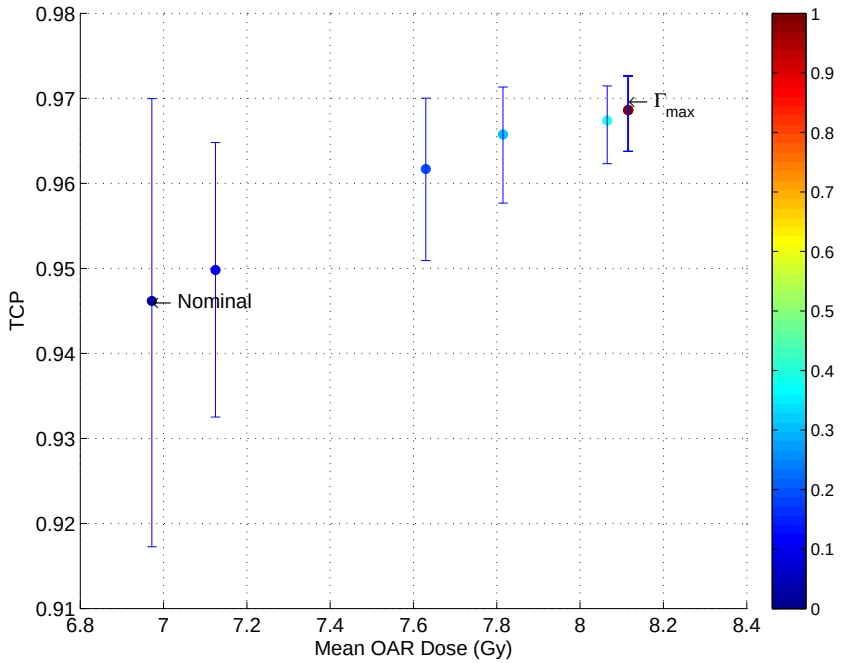
We can ascertain from Figure 3.9 that as we increase the value of Γ, the total sum of absolute deviations from prescription dose decrease. The effect is more pronounced in the worst case uncertainty scenario. Figure 3.10 shows that as we increase Γ, the maximum overdose increases. The effect is more pronounced in the uniform uncertainty scenario. Figure 3.11 shows that as we increase Γ, the maximum underdose decreases. The differences between the robust and nominal models are more apparent in the worst case uncertainty scenario. From Figure 3.12 we can observe that the objective function acts as a linear approximation for the nonlinear TCP function; they are inversely correlated. By limiting the underdose and the maximum underdose we maximize the overall TCP.

3.3 Sensitivity analysis

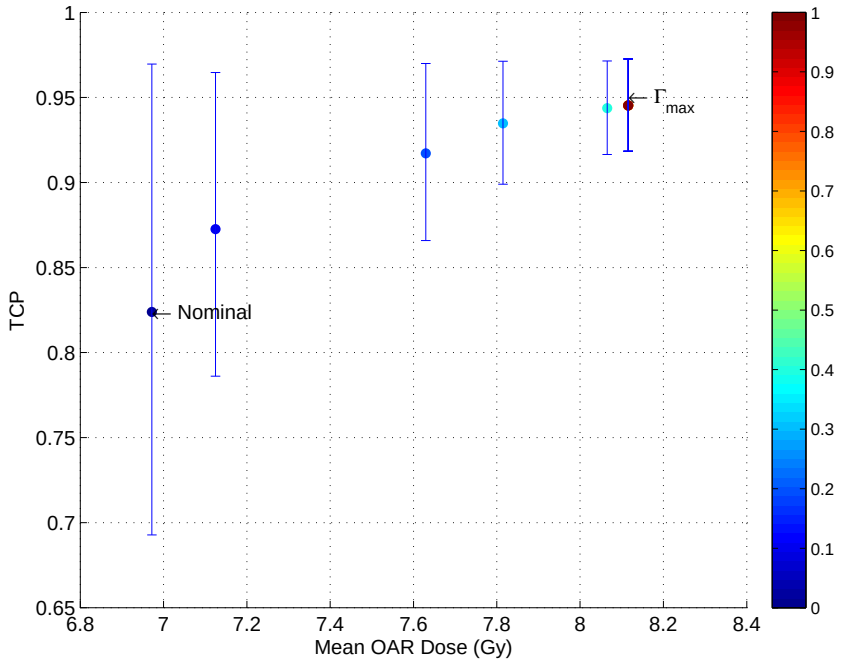
In this section we conduct a sensitivity analysis on several free parameters in our biological models and uncertainty interval. We investigate the effects of a simulated SUV being realized outside of the uncertainty interval, increased clonogen density, increased cell proliferation rate, and a steeper SUV to oxygen curve. In each analysis, we are mainly concerned with the effects of increased radioresistance and how much the previous findings degrade as a result.

Despite accounting for a certain length of uncertain interval, there is always the chance that the true SUV lies outside of the assumed range. We consider this situation by assuming that the length of the half-interval has been multiplied by a factor of $\epsilon > 1$. The true SUV of voxel i now lies in the range $[\theta_i - \epsilon\hat{\theta}_i, \theta_i + \epsilon\hat{\theta}_i]$. The treatment planning model is still solved using the range with half-interval $\hat{\theta}_i$. However during simulation, the realized SUV will be selected from the wider range with half-interval $\epsilon\hat{\theta}_i$. We set $\epsilon = 1.25$, which effectively means a half-interval length $\hat{\theta}_i = 0.75\theta_i$. Figure 3.13 and figure 3.14 compare the performance of the nominal and robust plans. Both plans experience increased degradation of performance. However, despite being optimized using parameters from the smaller original range, the robust plan experiences less degradation in performance than the nominal plan.

In our original formulation we assumed that the parameters T_p (potential cell doubling time) and

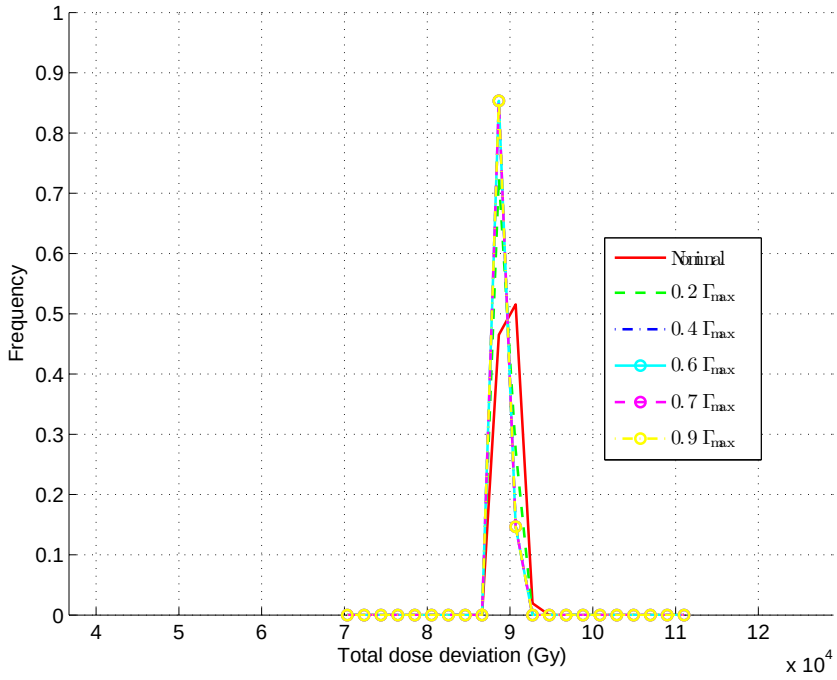


(a) Uniform uncertainty

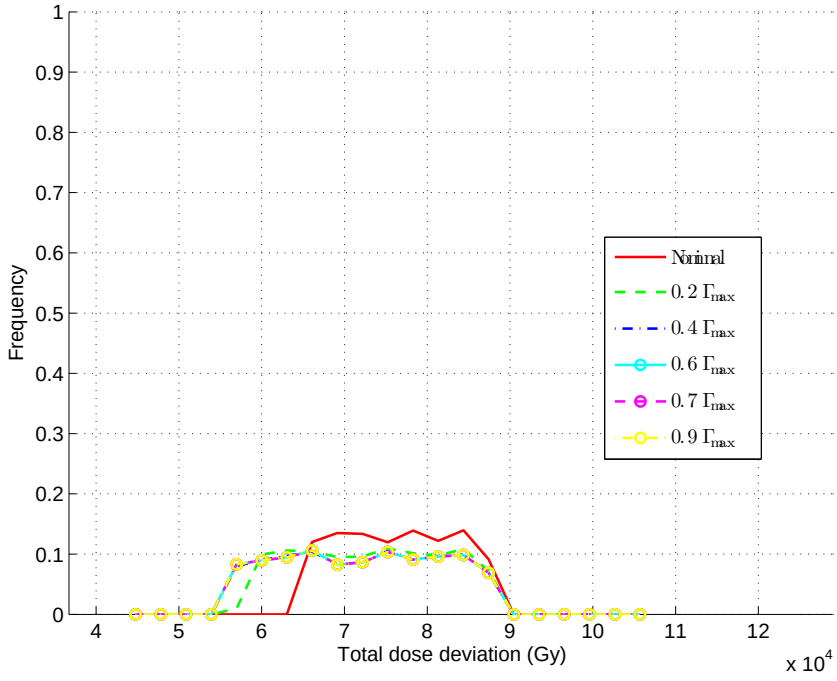


(b) Worst case uncertainty

Figure 3.8: Tradeoff curve between mean TCP and mean OAR dose. The color of the circle represents a unique Γ value. The circle itself is the mean TCP value for that Γ value. The blue lines represent the upper and lower bounds of the TCP for a particular Γ value.

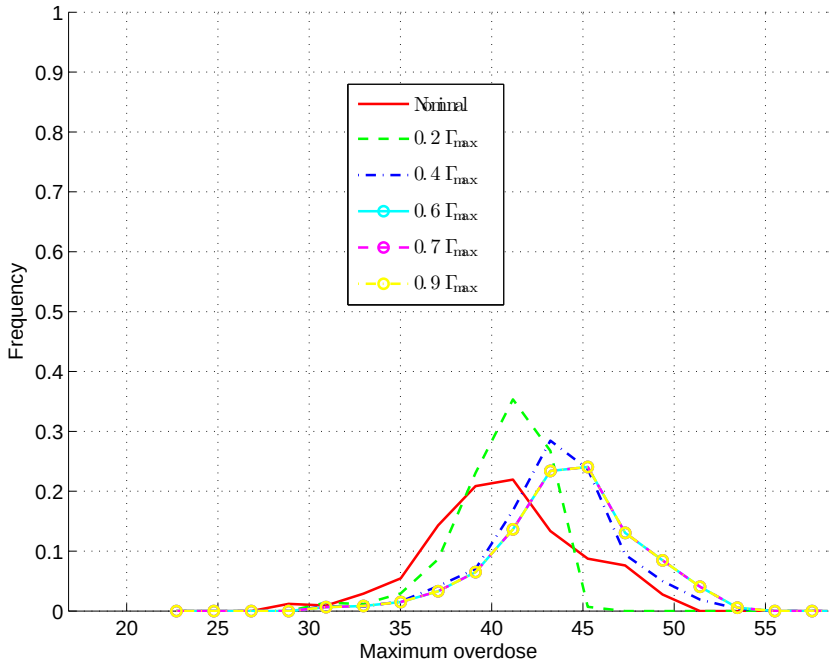


(a) Uniform uncertainty

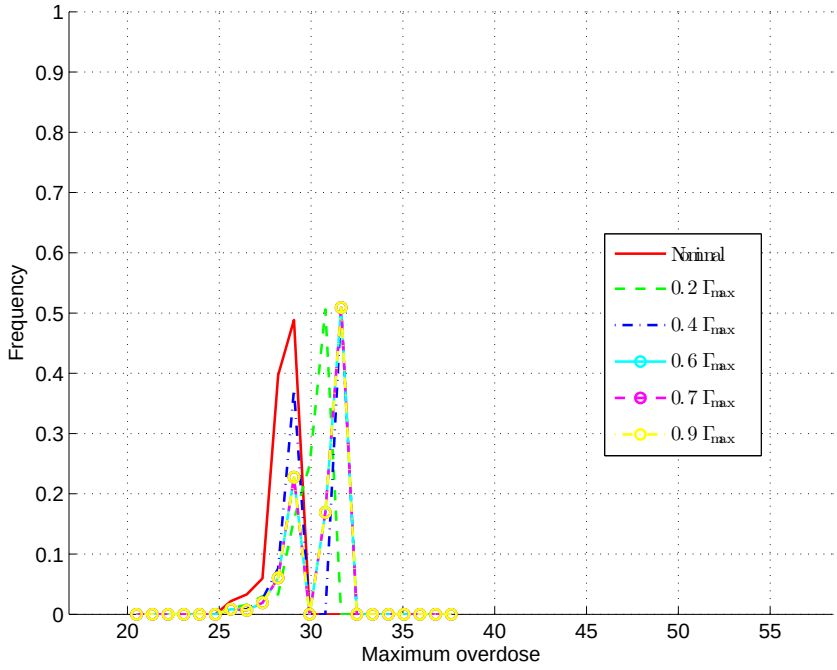


(b) Worst case uncertainty

Figure 3.9: Sum of total dose deviations (over and under) from prescription dose

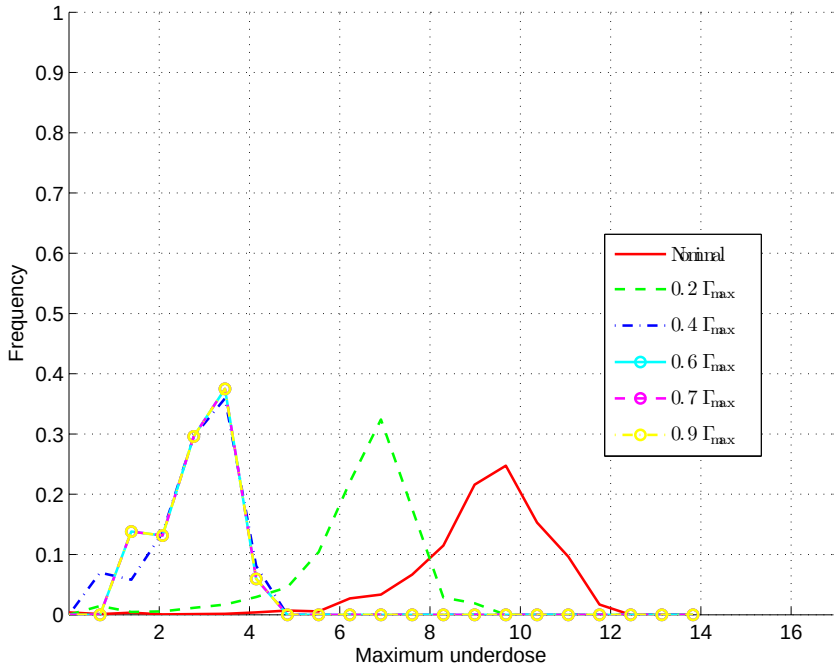


(a) Uniform uncertainty

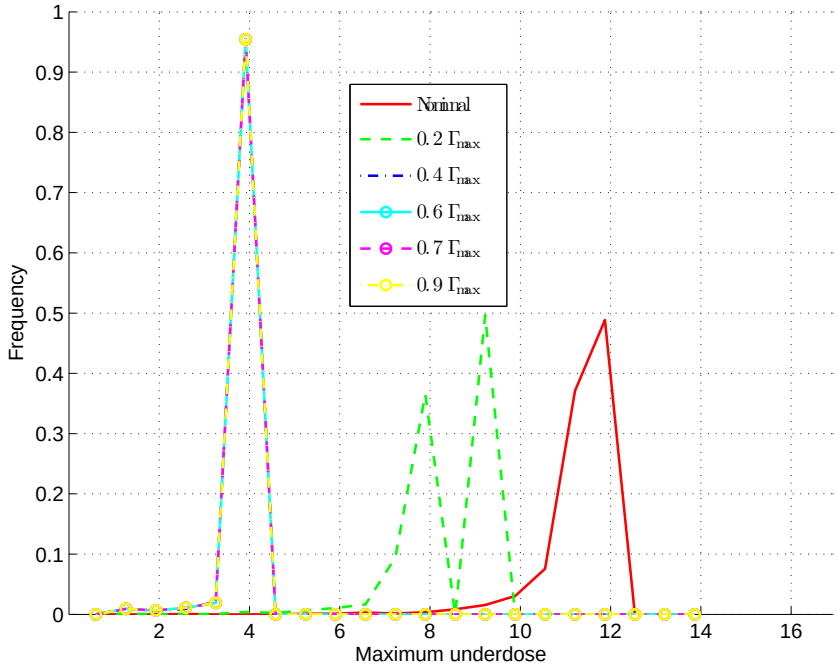


(b) Worst case uncertainty

Figure 3.10: Maximum overdose

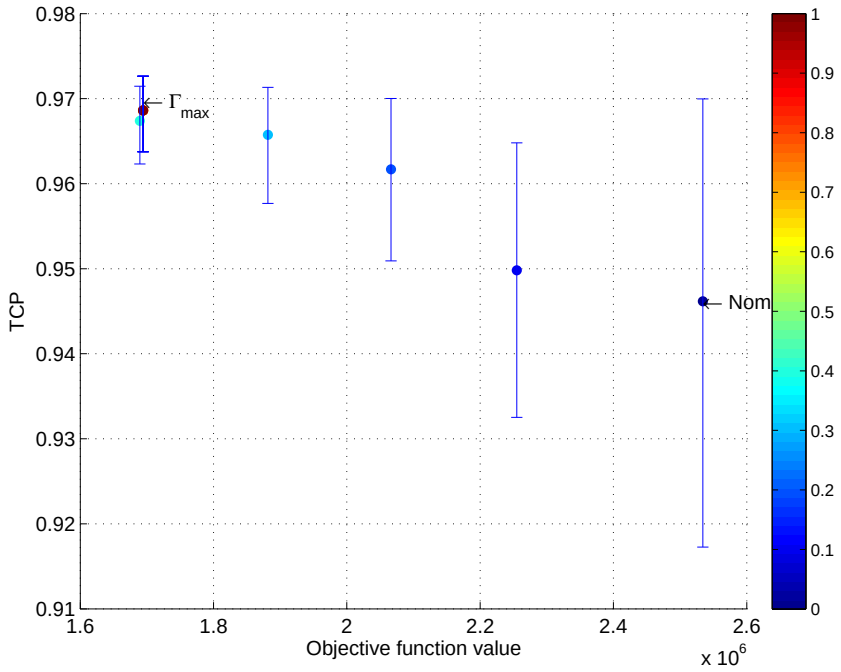


(a) Uniform uncertainty

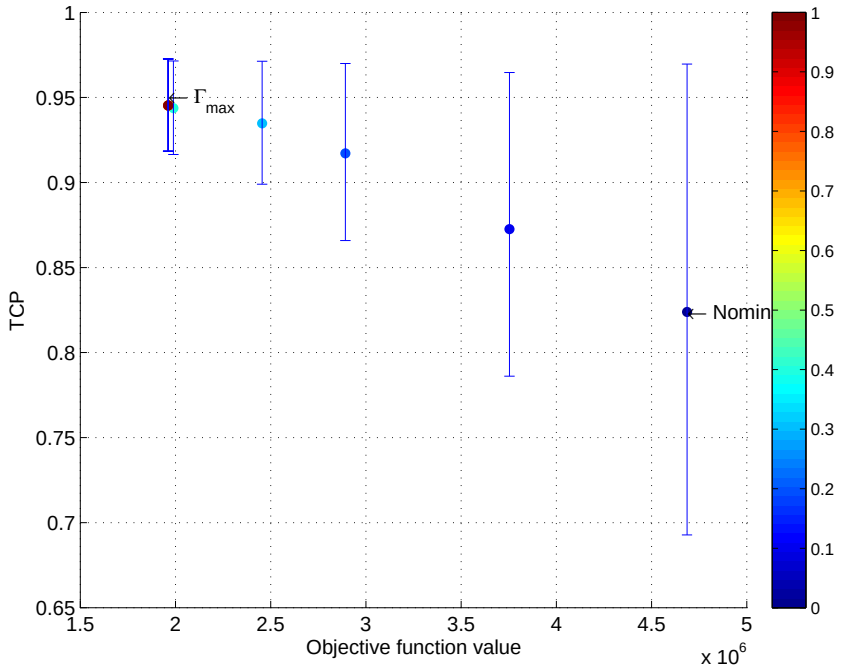


(b) Worst case uncertainty

Figure 3.11: Maximum underdose

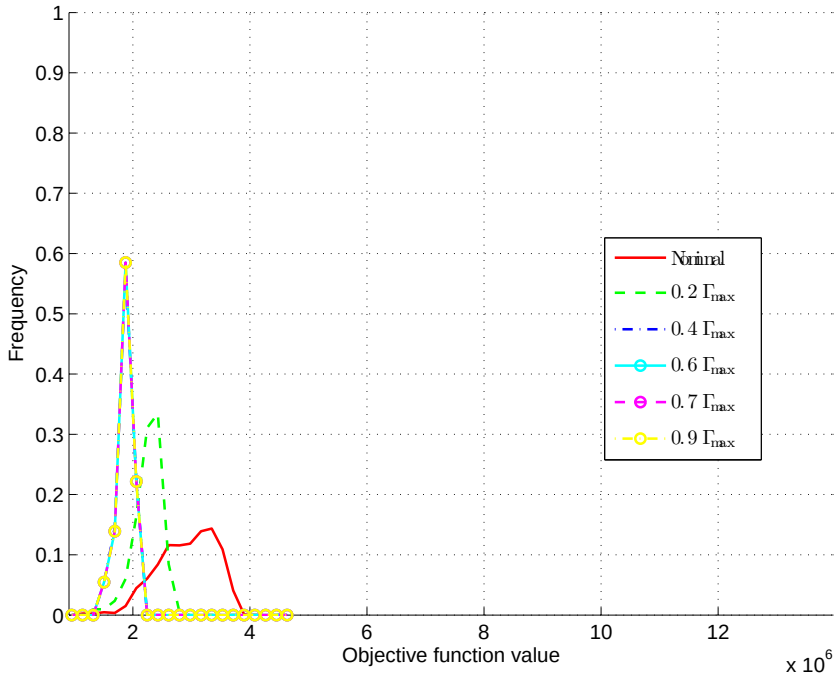


(a) Uniform uncertainty

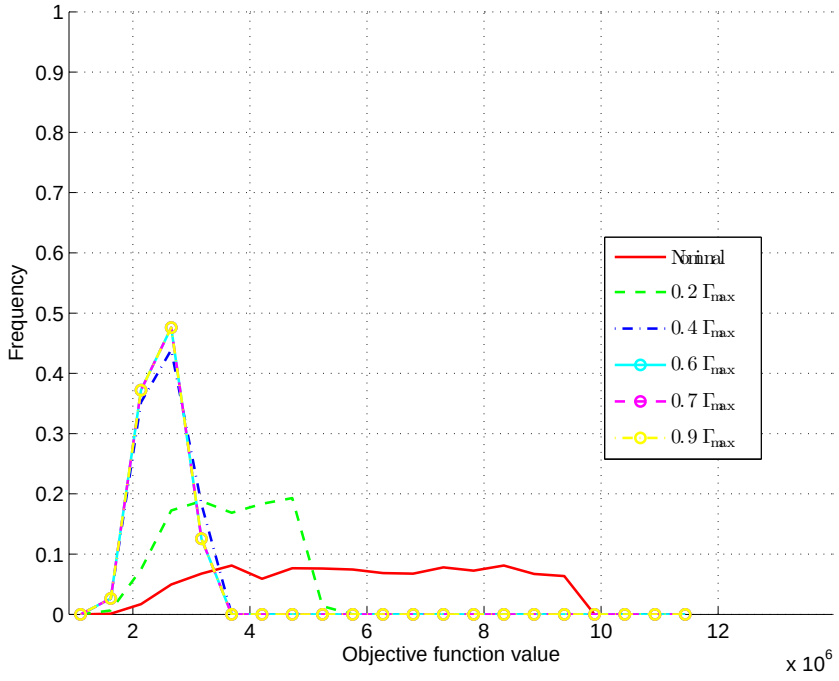


(b) Worst case uncertainty

Figure 3.12: Empirical relationship between objective function value and TCP. The color of the circle represents a unique Γ value. The circle itself is the mean TCP value for that Γ value. The blue lines represent the upper and lower bounds of the TCP for a particular Γ value.

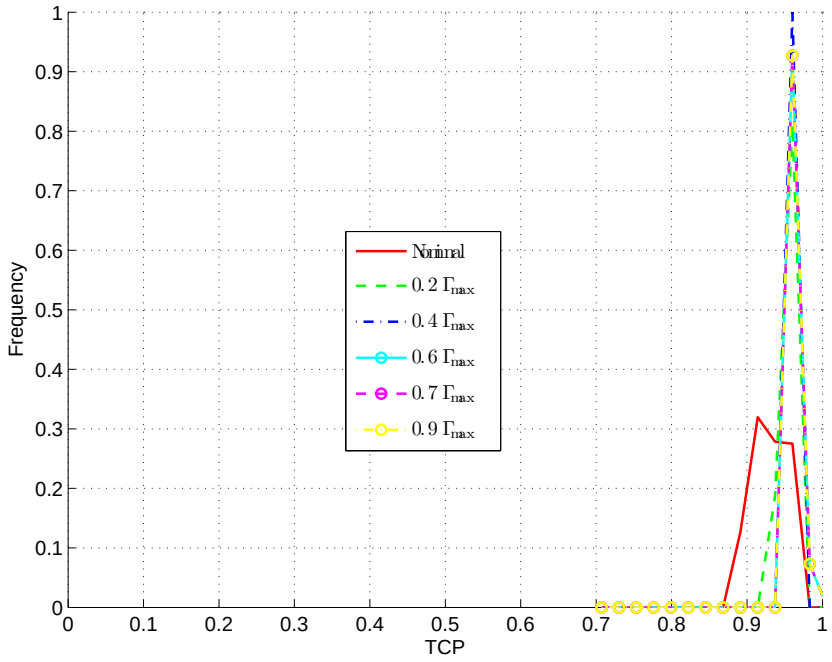


(a) Uniform uncertainty

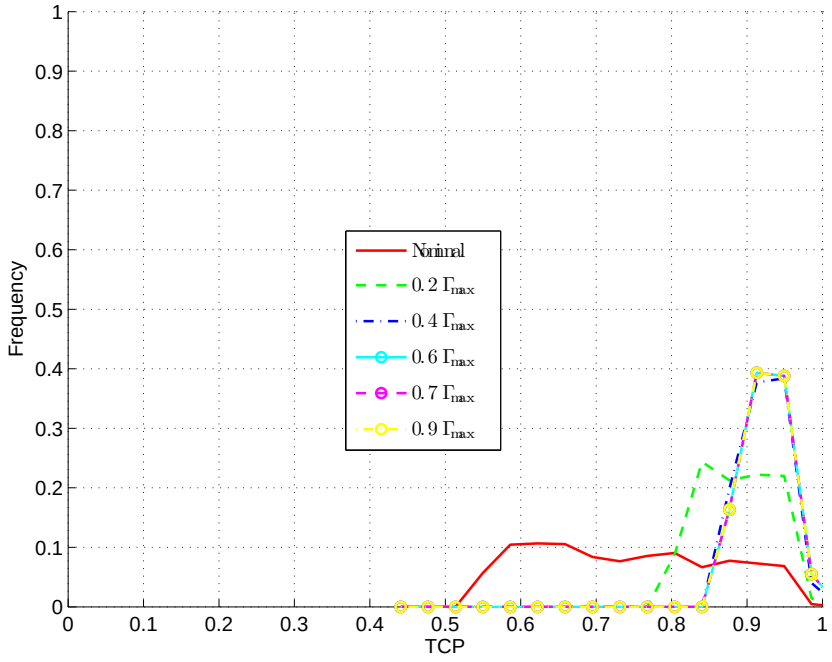


(b) Worst case uncertainty

Figure 3.13: Comparison of realized objective function values between nominal and robust plans, $\epsilon = 1.25$



(a) Uniform uncertainty



(b) Worst case uncertainty

Figure 3.14: Comparison of realized TCP values between nominal and robust plans, $\epsilon = 1.25$

ρ (clonogen density) were static with values of 3 days and 10^7 clonogens/cm^3, respectively. While we did not account for variation from these point values explicitly, we performed a sensitivity analysis to investigate the robustness of the models to their uncertainty.

From Figure 3.15, we can observe that the mean TCP of both the robust and nominal plans is reduced in the uniform and the worst case uncertainty scenarios when the clonogen density ρ is 25% higher than estimated. From Figure 3.16, we can observe that the mean TCP of both the robust and nominal plans is reduced in the uniform and the worst case uncertainty scenarios, when the potential cell doubling time T_p is 25% shorter. The sensitivity of the solutions to T_p is higher than ρ or ϵ. The robust does not degrade as much as the nominal. Being robust against one type of uncertainty, SUV, has provided some protection against uncertainty in other parameters. This is because regardless of the source of uncertainty, the downstream effect is all combined into the prescription dose. Since being robust against SUV uncertainty provides protection against an increase in the prescription dose, this would also provide some protection against a higher increase in prescription dose due to some other unconsidered source of uncertainty.

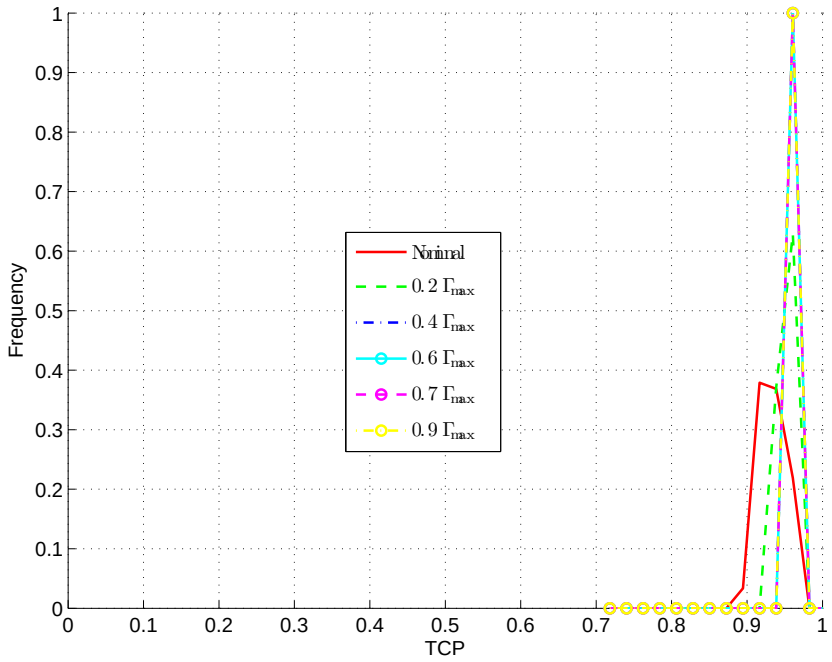
Since the relation between SUV and O_2 was merely theoretical, we also investigated the sensitivity of the plans against the uncertainty in the conversion function. We set the parameter ϕ_1 to be 0.75 of the original value of ϕ_1. We can observe in Figure 3.17 that the nominal model suffers a large degredation under both the uniform and worst case uncertainty scenarios and the robust model does not degrade as much as the nominal.

3.4 Discussion

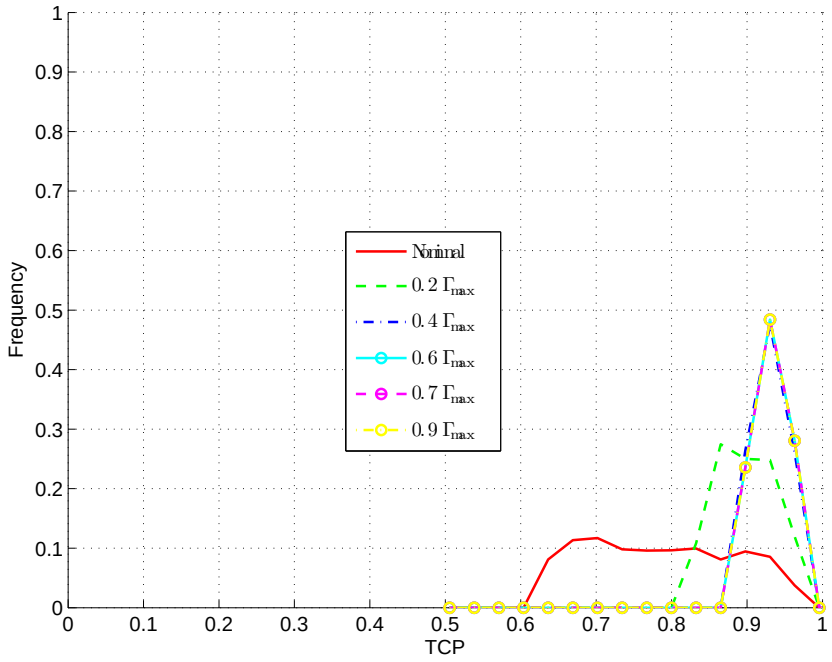
While the range of the robust plans have an improved lower bound over the nominal plan, the distribution of the nominal plan's TCP values under the uniform uncertainty scenario are similar to that of the robust plans in Figure 3.6a. It is only under the worst case uncertainty scenario in Figure 3.6b that the robust plan is more clearly differentiated from the nominal plan. Since under the uniform uncertainty scenario we expect SUV to be below their nominal values half the time, this should improve the TCP for a constant dose distribution.

Figure 3.17 we can observe that the resultant TCP values experience a much larger amount of degradation when the SUV to oxygen conversion function is perturbed. However, the robust model still performs relatively better than the nominal model in both the uniform and the worst case uncertainty scenarios. This high sensitivity to the precise mathematical form of the conversion function emphasizes the necessity of determining the relationship between a PET tracer and the required prescription dose. Comparing Figure 3.6b and Figure 3.17b under the worst case uncertainty scenarios, we can observe that there exists a robust solution which is somewhat comparable to the nominal model's performance under no uncertainty in its SUV to oxygen conversion function. This demonstrates the ability of the robust solution to deal with sources of uncertainty which were not even accounted for originally.

The quality of the DVH is highly dependent on disease presentation. In certain patient cases, it may be more difficult to redistribute dose to the target. The models shown in section 3.1.4 only considered mean and max doses to the OAR. We were able to meet DVH constraints by adjusting the penalty given to total OAR dose. Binary integer dose volume constraints were omitted to facilitate computational efficiency. Linear approximations of dose volume constraints, such as minimizing mean upper tail dose to OARs (CVaR) (Romeijn et al., 2003), could be included without increasing computational overhead

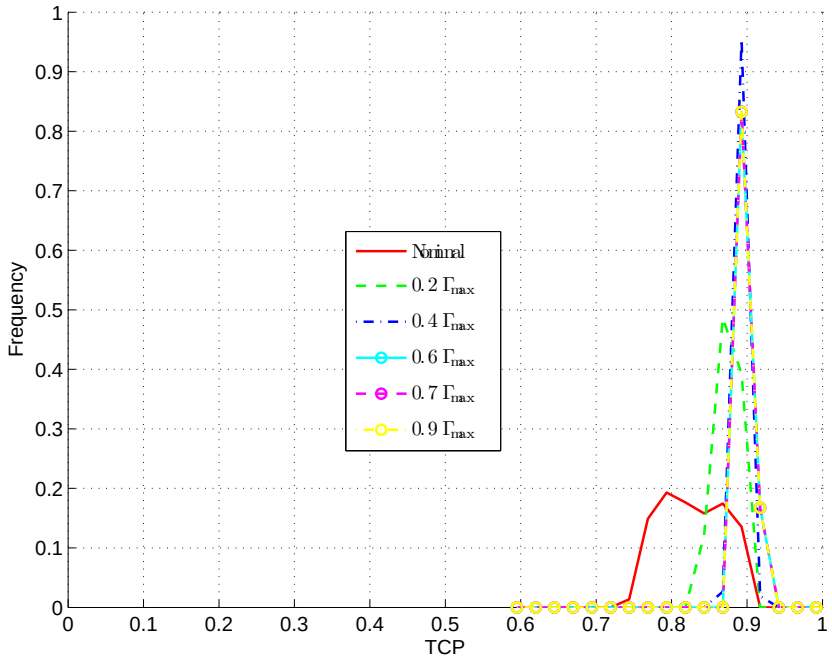


(a) Uniform uncertainty

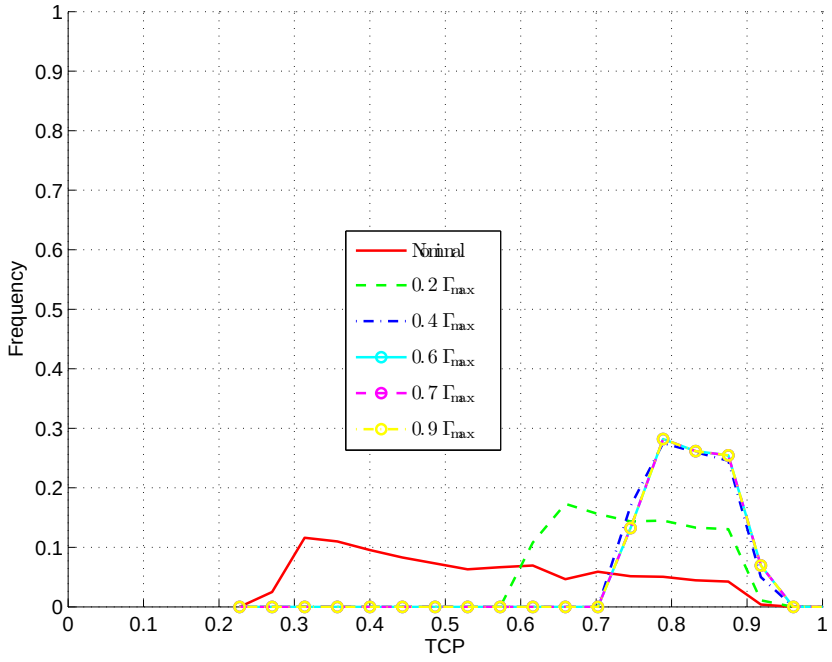


(b) Worst case uncertainty

Figure 3.15: Comparison of realized TCP values between nominal and robust plans, 1.25ρ

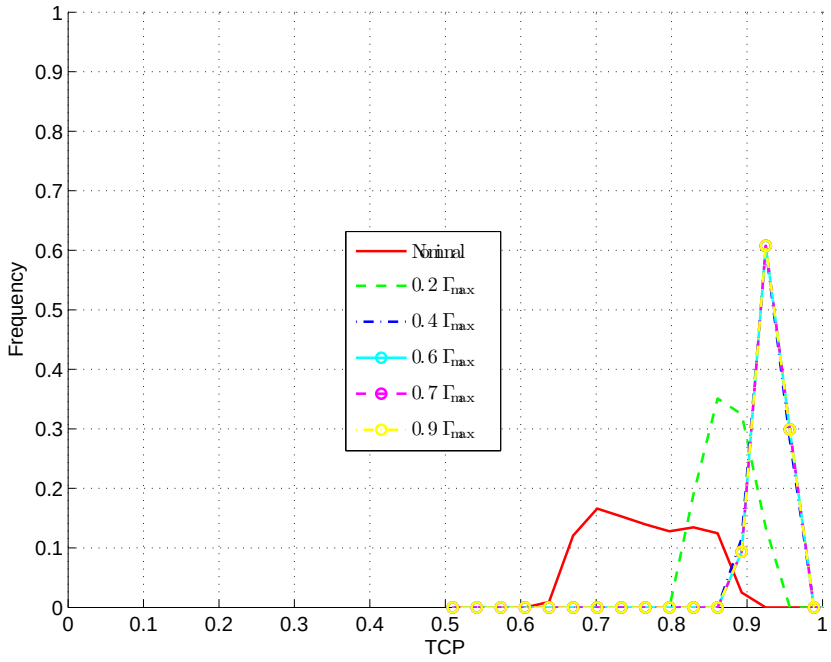


(a) Uniform uncertainty

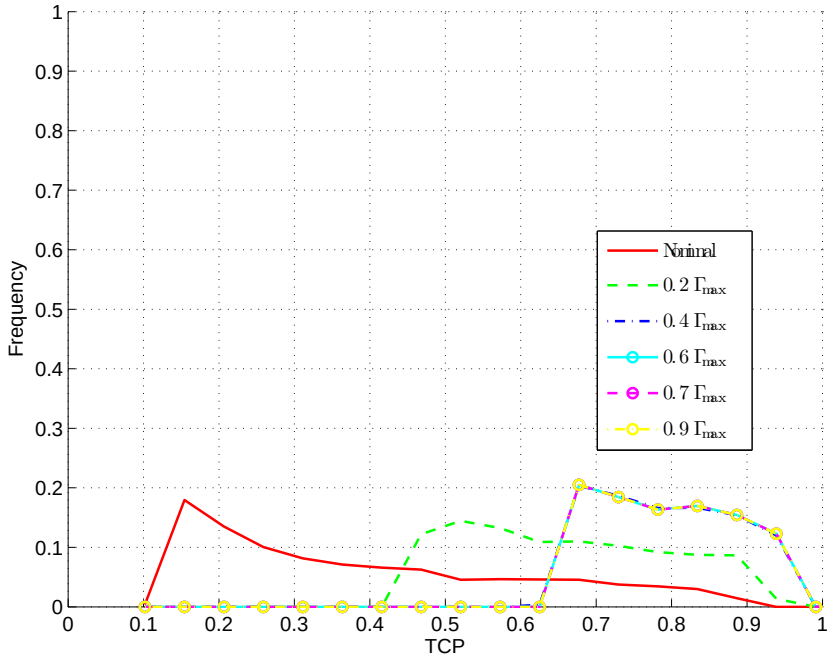


(b) Worst case uncertainty

Figure 3.16: Histogram of realized TCP values of nominal and robust plans, $0.75T_p$



(a) Uniform uncertainty



(b) Worst case uncertainty

Figure 3.17: Comparison of realized TCP values between nominal and robust plans, $0.75\phi_1$

by a substantial amount.

We can also observe the lack of change of the TCP histogram in Figure 3.6 with higher values of Γ. Since the CTV includes part of the non-PET avid regions to account for microscopic disease extension, some voxels have a nominal SUV as low as 0.0170. Even if this SUV doubled in value, it still would have no effect on the prescribed dose according to the conversion functions in Figure 3.2. The robust model prioritizes voxels with a high SUV, so the low SUV voxels are only incorporated at higher Γ values. If the thresholding nature of the conversion functions were changed to a curve with a continuous conversion from SUV to dose, we would expect a change to be present for all values of Γ, just that the change becomes incrementally smaller as Γ increases.

The objective function used in both the nominal and robust models was a weighted sum of multiple objective functions. The results are on the planner's prioritization of the different objectives. Usually, objective functions involving deviation from prescribed dose utilize quadratic penalties rather than linear ones. However, we instead chose to use a simpler absolute deviation to keep the problem linear and reduce computational load. Quadratic deviations from dose could also be modelled linearly using piecewise linear functions as an approximation. Another consequence of using a dose deviation penalty function (linear or nonlinear) is that it is only an approximation of the underlying TCP function. Therefore, an improvement in the objective function under uncertainty may not always translate to an exact improvement in the TCP value.

Despite there being no penalty on overdose, we can observe from Figure 3.9 that the total deviations from prescription dose, whether overdose or underdose are reduced in the robust model.

While this study considered the effects of uncertainty in the presentation of the PET image, we did not consider the effects of motion on the delivery of the dose. Delivery of a heterogeneous dose prescription in the presence of tumor motion can drastically alter the expected outcome of the planned treatment. Robust optimization has been applied to mitigate the effects of intrafraction tumor motion (Chan et al., 2006) and interfraction tumor motion (Chu et al., 2005; Olafsson and Wright, 2006) on a uniform tumor dose.

3.5 Partial volume effect simulation

The previous sections investigated more mathematically abstract realizations of uncertainty. In the uniform and worst case uncertainty scenarios a change in SUV could be realized anywhere in the CTV structure. We basically assumed that each voxel changes independently from each other. We will now consider an uncertainty scenario that is more biologically realistic. The PVE is a source of uncertainty which distorts the values observed in the PET scan. It actually refers to two separate imaging issues. The first is the finite spatial resolution of the imaging system, which introduces blurring into the image. Some of the SUV can be though of as "spilling out" into neighbouring regions, thereby changing the nominal number of clonogens. The second is image sampling or the tissue fraction effect. Since the image is discretized into voxels, some voxels will contain tissues from different ROIs and the resultant intensity is the mean of the intensities of the different tissues. Overall, PVE decreases the values observed in the image (Soret et al., 2007). PVE has been shown to have a maximum effect on the observed image of -30%, i.e. the true SUV is supposed to be 30% higher than observed near the edges (Boellaard, 2009).

We attempt to simulate the PVE as follows. We consider the PTV to be composed of 7 nested shells each of 4 mm thickness. We assume the outermost shell will have the highest uncertainty due

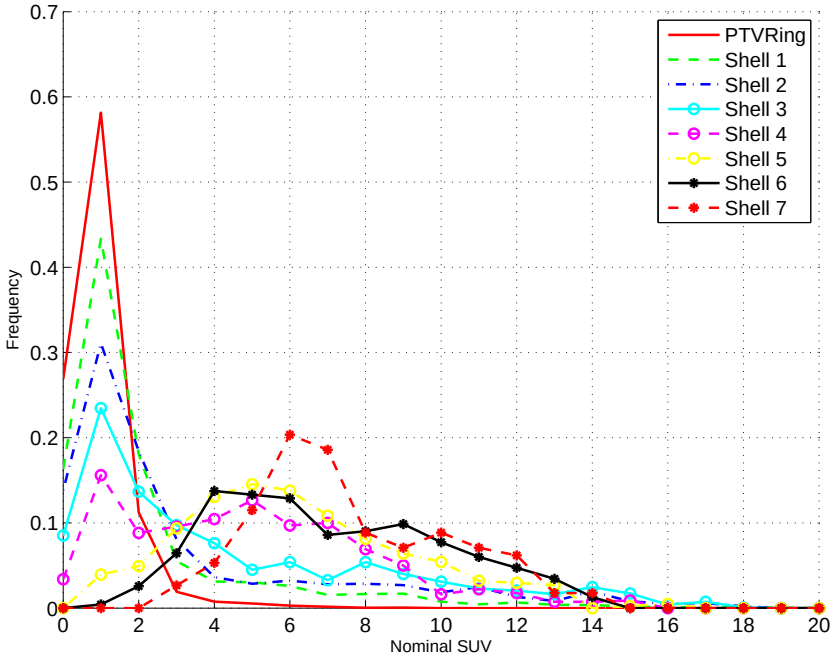
	Shell						
	1	2	3	4	5	6	7
Half-interval size	$0.3\theta_i$	$0.25\theta_i$	$0.2\theta_i$	$0.15\theta_i$	$0.1\theta_i$	$0.05\theta_i$	$0.05\theta_i$

Table 3.4: Size of the uncertainty half-interval for shells. Shell 1 is the outermost shell. Shell 7 is the innermost shell.

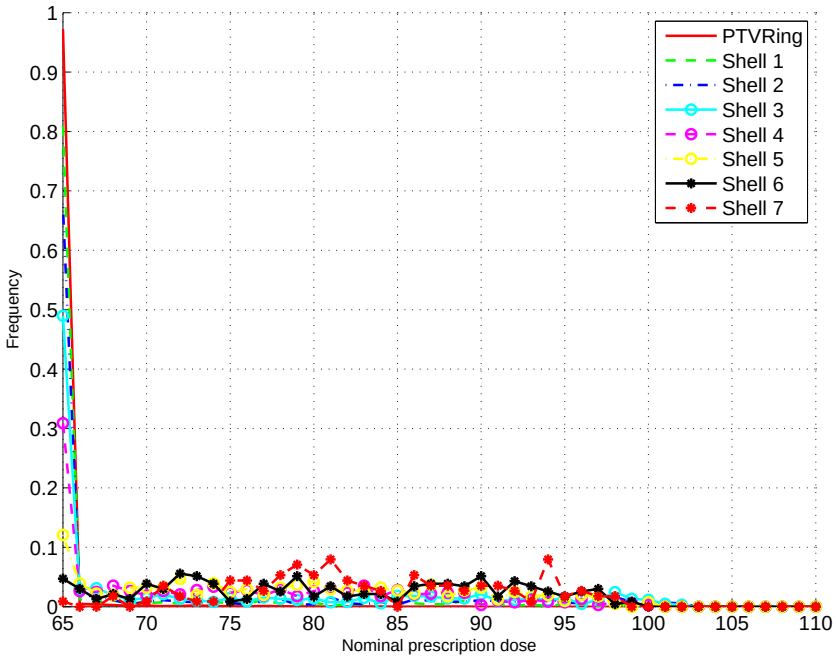
to its proximity to neighbouring non-target structures. The uncertainty decreases as we move from the outermost shell to the innermost shell. Table 3.4 outlines the size of the uncertainty half-interval. We only considered voxels potentially increasing in SUV. The simulation begins by choosing a random number of voxels to realize a change from their nominal SUV. The voxels on the outermost ring will be selected first until there are no more voxels in that ring from which to choose. Then the simulation will begin to select voxels on the next outermost ring to realize a change; this continues until the specified number of voxels to realize a change have been selected. Figure 3.18 shows the distributions of nominal SUV and prescription dose across the different shells in the PVE simulations. As we move from the outermost shell (Shell 1) to the innermost (Shell 7), we can observe an increase in the mean SUV and prescription dose.

From Figure 3.19 we can observe that the robust model still shows improvements in the mean and variance for both the objective function value and TCP. The simulations involving PVE used a maximum half-interval width of $0.3\theta_i$, while the abstract simulations used a half-interval width of $0.6\theta_i$. If we compare Figure 3.19a and Figure 3.6b, we can observe that the distribution of TCP has actually shifted to the left in the simulations involving PVE. This means that despite a smaller half-interval of uncertainty, the nominal and robust models performed relatively worse in the PVE simulations than in the abstract simulations. Based on the PVE simulation, areas on the outermost shell will have lower SUV but will experience the highest fractional deviation from their nominal SUV, while the areas on the inner shells will have higher SUV, but will experience proportionally less deviation. But a larger fraction of a smaller number could be potentially less than a small fraction of a larger number. Therefore, the model could still select the innermost voxels to protect against first, while the outermost voxels will experience a change first. At a high level, the protection of the robust model would effectively begin from inside and move outwards, while the realization of uncertainty would begin from the outside and move inwards. This inverted relationship between the robust model and the simulation could cause this degradation in performance.

From Figure 3.19a we can observe that lower values of Γ have a higher TCP than a higher value of Γ. A lower value of Γ having a better distribution of TCP initially seems inconsistent with the goal of the protection parameter. Intuitively, a higher protection parameter value should offer better TCP values in the face of uncertainty. However, the objective function is simply a linear approximation of the true objective which we are trying to optimize, TCP. The model is in fact, performing as expected as seen in Figure 3.19b, with increasing levels of protection resulting in improved realized objective function values despite the presence of uncertainty.

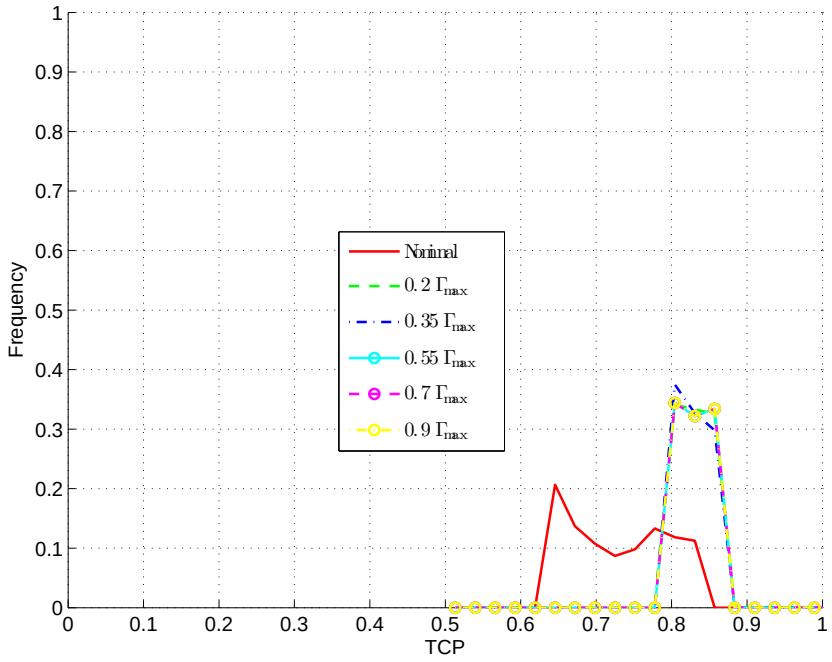


(a) Distribution of nominal SUV in PVE simulations

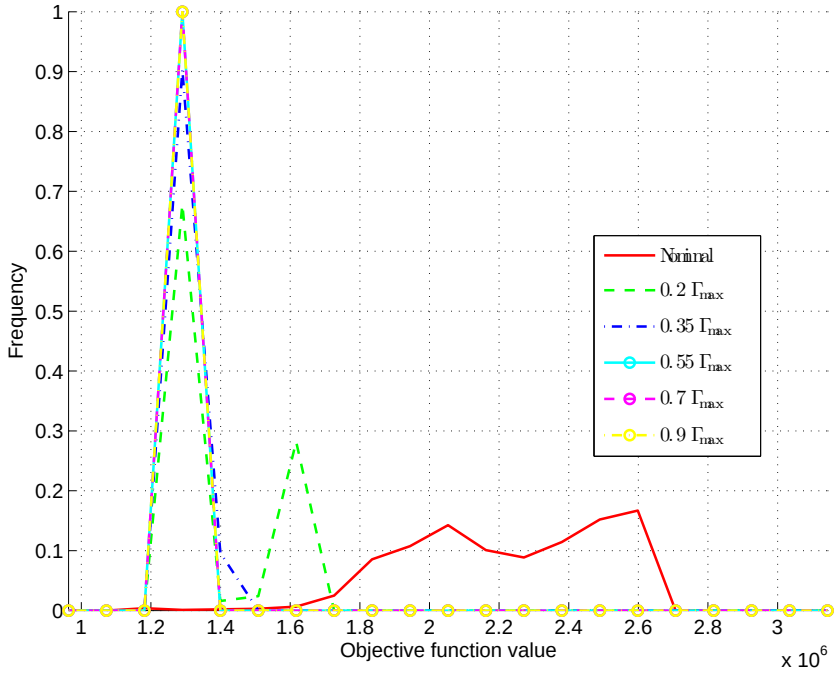


(b) Distribution of nominal prescription dose in PVE simulations

Figure 3.18: PVE simulation distributions. Shell 1 is the outermost shell and Shell 7 is the innermost shell.



(a) TCP under partial volume effect



(b) Objective function value under partial volume effect

Figure 3.19: Worst case uncertainty under partial volume effect

Chapter 4

Lung cancer response classification using pattern recognition

In this chapter, we apply machine learning techniques to predict patient response based on PET scans. The contributions of this chapter are as follows.

1. We compare the performance of different classifiers and feature selection and extraction methods using a PET dataset.

2. We investigate the usefulness of using a more detailed 4D gated PET/CT dataset.

This chapter is organized as follows. Section 4.1 describes the dataset, feature selection/extraction algorithms, and classifiers. Section 4.2 describes the results obtained from our classification experiments. Section 4.3 discusses the results.

4.1 Methods and Materials

This section describes the patient data, feature selection and extraction methods, and classifiers to be investigated.

4.1.1 Data

The data used in this study was provided by collaborators at Toronto's Princess Margaret Cancer Center. Raw 4D FDG-PET/CT images and tumour contours were exported from the hospital's clinical treatment planning system, Pinnacle, using the digital imaging and communications in medicine (DICOM) format. The data was then imported into the Computational Environment for Radiotherapy Research (CERR), a MATLAB-based research tool (Deasy et al., 2003). The PET scanner had a slice resolution of 3.9 x 3.9 mm / pixel and a slice thickness of 2 mm. Raw PET scanner values were converted to SUV.

The ROI in this study was taken to be the gross tumour volume (GTV). The primary GTV and nodal (spread of cancer to regional lymph nodes) GTV were treated as separate ROIs. A total of 62 tumors from 31 patients were analyzed in this study. A number of patients and tumours were removed due to inability to determine response and issues with contours. In the end, a total of 41 tumours were considered as the data set. From these 41 tumors, there were 36 that responded to treatment and 5 that

did not respond to treatment. We will refer to each of these 41 tumors as examples. At most 2 examples can come from a single patient.

Each site of primary and nodal disease was evaluated by a physician to determine response to radiotherapy. This classification was performed by a physician under the EORTC response criteria (Young et al., 1999). The EORTC criteria are as follows: complete metabolic response (CMR), partial metabolic response (PMR), progressive metabolic disease (PMD), and stable metabolic disease (SMD). A grouping was performed on the 4 EORTC response criteria to obtain a binary classification of response to treatment. The criteria CMR and PMR were considered as "response to treatment". The criteria PMD and SMD were considered as "non-response to treatment". This thresholding was performed due to the small sample size from each group, so that we could convert 4 small groups into 2 larger groups.

4.1.2 Features

Each patient used in this study had scans taken at 5 different intervals of the treatment process. The first was taken pre-treatment. The next 3 were taken during radiotherapy, with each set of scans being taken approximately 2 weeks apart. The final scan was then conducted approximately 3 months after radiotherapy was completed. Since the tumour is located in the lung, breathing motion results in a blurred image. Therefore, each set of 5 scans were taken at both exhale and inhale phases of breathing. However, only the features taken from exhale scans were used, since usually the majority of the patient's breathing is located at exhale and this results in less blurring of the PET image than an averaged scan. The following is a list describing the complete feature set.

- **IVH $\mathbf{V}_x$**: Interpretation of a time-series dataset of 3-dimensional images is difficult and would be prone to overfitting due to the high dimensionality, therefore the intensity volume histogram (IVH) is used to reduce the complexity of 3-dimensional image to a single curve. A sample IVH is shown in Figure 4.1. The x-axis is the SUV signal and the y-axis is the fractional volume. This curve represents the fractional volume of the region of interest (ROI) having an SUV of at least x% of $\mathrm{SUV}_{\max}$ (V_x). We discretized the IVH into 100 bins of length 1, so each bin covers 1% of the overall range. Therefore, the IVH data contains V_1 to V_{100}.
- **Max SUV**: The maximum SUV for that scan.
- **Mean SUV**: The mean SUV for that scan.
- **Median SUV**: The median SUV for that scan.
- **Peak SUV**: The 95^{th} percentile of SUV for that scan. This value should be less susceptible to image noise than the max SUV.
- **Volume**: Volume of the GTV for that scan.
- **V1.5SUV**: Volume with SUV greater than 1.5 for that scan.
- **V2SUV**: Volume with SUV greater than 2 for that scan.
- **V2.5SUV**: Volume with SUV greater than 2.5 for that scan.
- **V5SUV**: Volume with SUV greater than 5 for that scan.

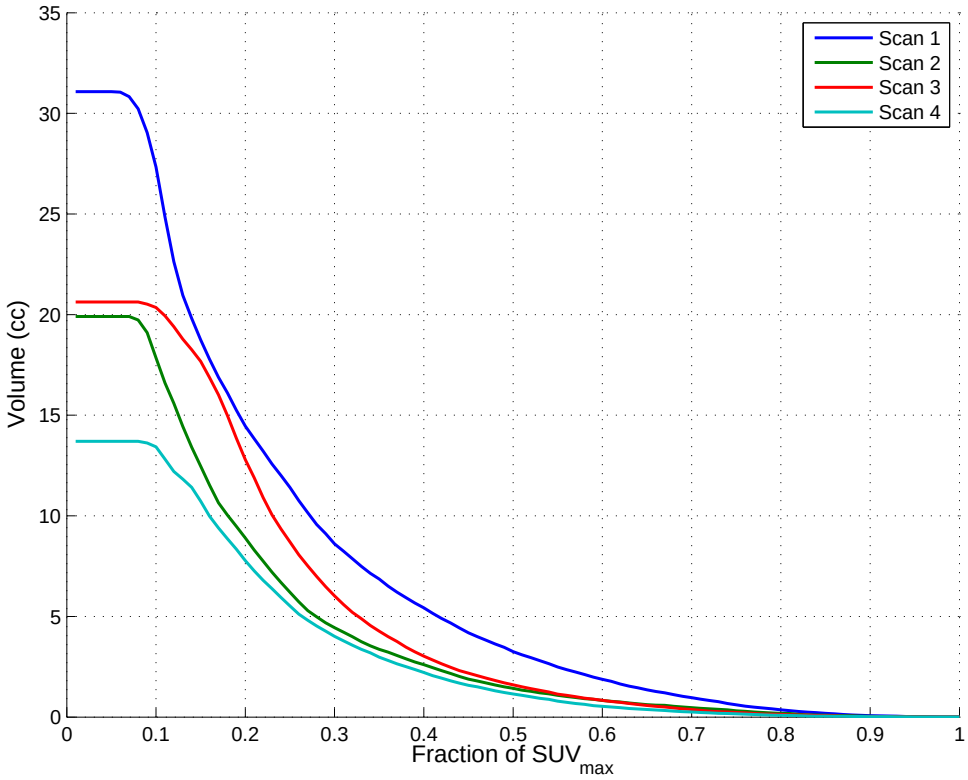


Figure 4.1: Intensity volume histograms (IVH) demonstrating the shrinking of metabolically active tumor volume over the course of treatment

- **V3SUV**: Volume with SUV greater than 3 for that scan.
- **V50pSUV**: Volume with SUV greater than 50% of the Scan 1 peak SUV.
- **Max50pSUV**: Volume with SUV greater than 50% of the Scan 1 max SUV.
- **V50pWSUV**: Volume with SUV greater than 50% of the peak SUV for that scan.

The original number of features for all 5 scans combined was 565. However, this only provided us with 18 examples since for a number of patients the final scan data was unavailable. Therefore it was decided to ignore the final scan as a trade off for incorporating more examples into the already small dataset. The final number of features for 4 scans was 452. This resulted in 41 examples. Intuitively, this should not impact the results too much, since the benefits of predicting patient response are realized the earlier you can make an accurate prediction. Making a prediction based on the subsequent scans leaves less room for any adaptive measures to be taken.

4.1.3 Preprocessing

Prior to actually training any classifier on the feature set, the data was preprocessed in order to minimize the effects of large absolute values of any one feature on the learning algorithm. This consisted of standardizing the data by subtracting the mean of each feature and dividing by its variance. The standardized data was then scaled to be within the range $[0, 1]$ for each feature.

4.1.4 Mathematical Notation

Here we introduce the notation which will be used in the sections hereafter. An example refers to a tumor, whether it be a primary or nodal tumor from any patient. At most 2 examples can come from 1 patient. Let each example be indexed by $i \in M$. Let each feature (dimension of each example) be indexed by $j \in N$. Let $\mathbf{x}_i$ be the feature vector for example i. Let x_{ij} be feature j of example i and y_i be the class of example i. Let y_i be the label of example i. A label of 1 corresponds to a responder and a label of -1 corresponds to a non-responder.

4.1.5 Feature selection and extraction

Due to the large dimensionality of the feature space in comparison to the number of data points available, we investigate a number of methods of selecting relevant features or extracting transformed features from the original features space to construct subsets of features which are more conducive to generalization (Guyon, 2006). We compare 3 different methods: feature selection through Spearman's rank correlation coefficient, recursive feature elimination, and principal components analysis.

Spearman's rank correlation coefficient

Spearman's rank correlation coefficient ρ is the Pearson correlation coefficient between the ranked variables. ρ_j for a given feature j is calculated as follows (Corder and Foreman, 2009).

$$\rho_j = \frac{\sum_{i \in M}(x_{ij}^r - \bar{x}_j^r)(y_i^r - \bar{y}^r)}{\sqrt{\sum_{i \in M}(x_{ij}^r - \bar{x}_j^r)^2 \sum_{i \in M}(y_i^r - \bar{y}^r)^2}} \tag{4.1.1}$$

The raw values of $\mathbf{x}_j$ and y_i are given a rank x_{ij}^r and y_i^r based on their position in ascending order. Rather than a linear assumption of the relationship between variables, the Spearman rank correlation coefficient is a nonparametric measure and only models the relationship as a monotonic function. We use the Spearman rank corelation coefficient and its p-value in a univariate feature selection method described in Algorithm 4.1. This method evaluates each feature independently of others and does not consider interaction between features.

Algorithm 4.1 Univariate feature selection algorithm

Require: Features and labels for each example
1: **for** $feature$ in $data$ **do**
2: Calculate its Spearman ρ and $p-$value
3: **end for**
4: **for** $feature$ in $data$ **do**
5: Add feature into feature set with next highest $|\rho_j|$
6: Train a classifier on feature set
7: Conduct LOOCV using trained classifier
8: **end for**
9: Select feature subset with the highest LOOCV accuracy
10: **if** Multiple feature subsets are tied **then**
11: Select feature subset with smallest number of features
12: **end if**
13: **return** Best feature subset

Recursive feature elimination

Recursive feature elimination utilizes the linear SVM shown in Equation 4.1.2 to fit an initial model and removes features with the smallest absolute weight value. After each model has been fitted, its cross-validation error is measured.

$$\min_{\mathbf{w}} \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i\in M}\max(0, 1 - y_i\mathbf{w}^T\mathbf{x}_i)^2 \tag{4.1.2}$$

A backwards elimination search was performed over the feature space rather than stopping when the cross-validation accuracy decreased. The feature subset with the highest cross-validation accuracy was selected. If multiple subsets of features were tied for highest cross-validation accuracy, then the subset with the smallest number of features was chosen.

Algorithm 4.2 RFE Algorithm

Require: Features and labels for each example
1: **for** $feature$ in *data* **do**
2: Train a linear SVM classifier on feature set
3: Conduct LOOCV using trained classifier
4: Remove feature from feature set with the lowest weight w_j
5: **end for**
6: Select feature subset with the highest LOOCV accuracy
7: **if** Multiple feature subsets are tied **then**
8: Select feature subset with smallest number of features
9: **end if**
10: **return** Best feature subset

Table 4.1 summarizes the list of features, their Spearman's rank correlation coefficient, and p-value. Here we omit showing V_x values not divisible by 10 for the sake of brevity.

PCA

Principal components analysis (PCA) is used as a method to reduce the dimensionality of the feature space (Jolliffe, 1986). It attempts to represent the high dimensional space D as a linear projection onto C lower dimensions, where $C << D$. Each of these dimensions in C are orthogonal to each other and are known as the principal components. PCA has the property that the principal components are ordered so that the first principal component explains the maximum variance in the original data, with each subsequent orthogonal principal component accounting for less and less variance in the data. Algorithm 4.3 (Smith, 2002) describes the integration of PCA within a feature extractor.

4.1.6 Unbalanced class distribution

The fraction of non-responding examples in the dataset was only $\frac{5}{41} = 0.12$. In order to compensate for a small class size of non-responders, this class was oversampled to obtain a larger minority class distribution. Oversampling of the minority class is one of the most common methods to deal with an underrepresented class size (Van Hulse et al., 2007).

	Spearman's ρ				p-value			
Feature	Scan 1	Scan 2	Scan 3	Scan 4	Scan 1	Scan 2	Scan 3	Scan 4
V_{10}	-0.0315	-0.0189	-0.0189	-0.0756	0.8450	0.9066	0.9066	0.6385
V_{20}	-0.0693	-0.0504	-0.0504	-0.1575	0.6668	0.7544	0.7544	0.3254
V_{30}	-0.1260	-0.1197	-0.0756	-0.2331	0.4325	0.4560	0.6385	0.1425
V_{40}	-0.1953	-0.1701	-0.1008	-0.1701	0.2211	0.2877	0.5306	0.2877
V_{50}	-0.2331	-0.2079	-0.1638	-0.1197	0.1425	0.1922	0.3062	0.4560
V_{60}	-0.2457	-0.2520	-0.1764	-0.1323	0.1216	0.1120	0.2700	0.4097
V_{70}	-0.2457	-0.2205	-0.2079	-0.0945	0.1216	0.1660	0.1922	0.5568
V_{80}	-0.2142	-0.1953	-0.1764	-0.0756	0.1787	0.2211	0.2700	0.6385
V_{90}	-0.1890	-0.0882	-0.1953	-0.0567	0.2367	0.5835	0.2211	0.7248
Max SUV	-0.1512	-0.2142	-0.0945	0.1071	0.3454	0.1787	0.5568	0.5051
Mean SUV	-0.2205	-0.2709	-0.2520	-0.1197	0.1660	0.0867	0.1120	0.4560
Median SUV	-0.2079	-0.2457	-0.1575	-0.0693	0.1922	0.1216	0.3254	0.6668
Peak SUV	-0.1638	-0.2394	-0.1890	-0.0126	0.3062	0.1317	0.2367	0.9377
Volume	-0.0063	-0.0378	-0.0063	-0.0630	0.9688	0.8145	0.9688	0.6956
V1p5SUV	-0.0756	-0.1354	-0.1386	-0.1764	0.6385	0.3985	0.3875	0.2700
V2SUV	-0.0945	-0.1858	-0.1701	-0.1165	0.5568	0.2447	0.2877	0.4681
V2p5SUV	-0.1197	-0.1921	-0.1796	-0.0536	0.4560	0.2288	0.2611	0.7391
V5SUV	-0.1733	-0.2625	-0.2500	0.0931	0.2785	0.0973	0.1149	0.5626
V3SUV	-0.1386	-0.2110	-0.1900	-0.0159	0.3875	0.1853	0.2341	0.9216
V50pSUV	-0.2331	-0.2902	-0.0948	0.1438	0.1425	0.0657	0.5553	0.3696
Max50pSUV	-0.1512	-0.2142	-0.0945	0.1071	0.3454	0.1787	0.5568	0.5051
V50pWSUV	-0.2331	-0.2079	-0.1638	-0.1197	0.1425	0.1922	0.3062	0.4560

Table 4.1: Features and their Spearman rank correlation coefficient and p-value

Algorithm 4.3 PCA Algorithm

Require: Features and labels for each example
1: **for** $feature$ in *data* **do**
2: For each feature $\mathbf{x}_j$ we subtract its mean $\bar{x}_j$.
3: **end for**
4: Calculate the empirical covariance matrix of $\mathbf{x}$
5: Find the eigenvectors and eigenvalues of the covariance matrix.
6: Order the eigenvectors by the largest eigenvalue.
7: **for** $feature$ in *data* **do**
8: Project the original onto the C eigenvectors with the C highest eigenvalues.
9: Train a classifier on the projected data
10: Conduct LOOCV using trained classifier
11: Increment C
12: **end for**
13: Select projected feature dimension C with the highest LOOCV accuracy
14: **if** Multiple projections are tied **then**
15: Select smallest projected feature dimension C
16: **end if**
17: **return** Best projected features for each example

4.1.7 Bootstrap

Bootstrap aggregation (bagging) involves creating new sets of training data by sampling with replacement from the original training data. With each set of training data, we train a classifier on it and make a prediction on the testing data. We take all the predictions from the classifiers and take the class which received the majority vote. In essence, we introduce a degree of randomness into the training data, which has been shown to improve the classification accuracy of predictors when small perturbations in the training set (Breiman, 1996). The randomness increases the diversity of the ensemble of classifiers. However, there are tradeoffs between accuracy and diversity (Windeatt, 2008). Given the small size of the dataset, changes in the dataset could result in large changes in the resulting classifier. We investigated the effect of the size of the bootstrap training data on the cross-validation accuracy.

4.1.8 Classifiers

Logistic regression (LR)

The logistic function is defined as

$$f(z) = \frac{1}{1 + e^{-z}} \tag{4.1.3}$$

and only takes on values in the range $[0, 1]$. In order to utilize it in a classification algorithm, we replace z with:

$$\mathbf{w}^T\mathbf{x} + b. \tag{4.1.4}$$

One can interpret w_j as the weight of each feature x_j. b is the initial bias term not related to any feature. To fit the function to the training data, we minimize the log likelihood as our loss function. Both L1 and L2 regularized models were investigated; their loss functions are described in Equation 4.1.5 and Equation 4.1.6, respectively:

$$\text{L1: } \min_{\mathbf{w}} \sum_j |w_j| + C \sum_{i \in M} \log(1 + \exp(-y_i \mathbf{w}^T \mathbf{x}_i)) \tag{4.1.5}$$

$$\text{L2: } \min_{\mathbf{w}} \frac{1}{2}\mathbf{w}^T\mathbf{w} + C \sum_{i \in M} \log(1 + \exp(-y_i \mathbf{w}^T \mathbf{x}_i)) \tag{4.1.6}$$

where w_j is the learned weight of feature j, C is penalty parameter, $\mathbf{x}_i$ is the feature vector for example i, and y_i is class of example i. L1 regularization tends to set a lot of weights to zero when compared to L2 regularization. The LIBLINEAR library for MATLAB (Fan et al., 2008) was used to train and test the logistic regression model.

Support vector machine (SVM)

Support vector machines find the linear separating hyperplane that provides the maximum margin separation between 2 sets of data. Model 4.1.7 assumes that a linear separating hyperplane does exist. However, if such a hyperplane does not exist, then Model 4.1.7 will be infeasible.

$$\begin{aligned} \min_{\mathbf{w},b,\xi} \quad & \frac{1}{2}\mathbf{w}^T\mathbf{w} \\ \text{s.t.} \quad & y_i(\mathbf{w}^T\mathbf{x_i} + b) \geq 1, \\ & \xi_i \geq 0. \end{aligned} \tag{4.1.7}$$

To allow for the possibility of an imperfect (perhaps due to mislabelled data, or extreme outliers) linear separating hyperplane we can use Model 4.1.8 (Cortes and Vapnik, 1995).

$$\begin{aligned} \min_{\mathbf{w},b,\xi} \quad & \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i\in M}\xi_i \\ \text{s.t.} \quad & y_i(\mathbf{w}^T\mathbf{x_i} + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0. \end{aligned} \tag{4.1.8}$$

ξ_i is a slack variable which allows for incorrectly classified examples. C is the penalty associated with incorrectly classifying the training data.

However, a linear separation in the input space may not be the best representation of the different classes. In order to obtain non-linear separation between the classes, we use the kernel trick (Boser et al., 1992). This will transform the original feature space and we fit a linear separating hyperplane in a higher dimensional space. The resulting separation in the original feature space can then be a nonlinear separation. The LIBSVM library for MATLAB Chang and Lin (2011) was used to train and test the SVM model.

4.1.9 Algorithm

Since the dataset is so small, we perform leave-one-out cross-validation in order to train our classifiers on as many data points as possible (Bishop and Nasrabadi, 2006). Algorithm 4.4 describes the overall algorithm we used.

Algorithm 4.4 Overall leave-one-out cross-validation algorithm

Require: Features and labels for each example
for *example* in *data* **do**
  Set aside *example* as testing data *test*
  Oversample minority class
  Select relevant features or transform features
  for number of bootstraps **do**
    Sample with replacement from training data to get bootstrapped training data
    Train a classifier on bootstrapped training data
    Predict class of *test*
  end for
  Average predictions from each classifier trained on their bootstrapped data
end for
Average accuracy rate on each *example* to obtain LOOCV accuracy
Take features were selected by the most classifiers
return LOOCV for classifier

Classifier	Num Scans	Maj./Min.	Num bootstrap	Feat. selection	CV Acc.
L2-reg LR	-	-	-	-	0.878
L1-reg LR	-	-	-	-	0.878
Linear SVM	1	3	10	Spearman	0.9268
Radial basis kernel SVM	2	3	80	Spearman	0.9024

Table 4.2: Best identified hyperparameter settings for a given classifier. Dashes in the settings represent classifiers which were unable to outperform the naive classifier; many different setting combinations resulted in the same accuracy. Maj./Min. represents the fraction of majority class to minority class.

4.2 Results

Consider a naive classifier that only bases its decision on the majority class within the training set. This naive classifer would obtain a CV accuracy of $\frac{36}{41} = 0.878$. Any classification algorithm needs to exceed this CV accuracy in order to be of any use. Appendix A contains a complete list of all combinations of experiments performed. Here we present only the results which were able to outperform the naive classifier for the sake of brevity.

The L1-regularized and L2-regularized logistic regression classifiers were unable to outperform the naive classifier; at most they matched the naive classifier's performance. The logistic regression based classifiers appeared to be sensitive to the degree of minority oversampling.

The SVM-based classifiers were able to outperform the naive classifer. The linear SVM was able to achieve a maximum CV accuracy of 0.9268, while the SVM using the radial basis function as its kernel function was able to achieve a maximum CV accuracy of 0.9024. The CV accuracy of 0.9268 for the linear SVM is actually the accuracy resulting from inverting the classifications made by the linear SVM. The original linear SVM classifer was actually a poor performer, but by inverting its decisions we can construct a classifier that performs better than the other classifiers. This inversion creates a good classifier because this is only a binary classification problem; if a classifier predicted every test case as the opposite of the true label, inverting the prediction should provide perfect accuracy. We can extend this logic to classifiers that perform poorly to construct better classifiers. The SVM-based classifiers were more stable across different degrees of oversampling. Table 4.2 shows the hyperparameter settings used to obtain the highest CV accuracy.

From Figure 4.2 we can observe that incorporating more bootstrap samples appeared to improve performance. Using only a single training set resulted in generally reduced performance.

Using PCA as a feature extraction method did not improve the performance of any of the classifiers. RFE and the Spearman's rank correlation-based feature selection methods were able to select feature subsets that performed better than the naive classifier. The majority of instances where the trained classifier performed better than the naive case utilized the Spearman rank correlation based feature selection method.

Since the feature selection algorithm is embedded within each training-testing split of the LOOCV, each split potentially has a different set of features selected. It is in clinical interests to identify a set of features which are predictive of patient response. Since the feature subsets are different in each split, we adopt a frequency approach to feature selection similar to Baek et al. (2009). Feature selection through this type of frequency approach has been shown to result in more stable feature sets. We select the features based on their frequency occurrence across all different training-testing splits of the dataset.

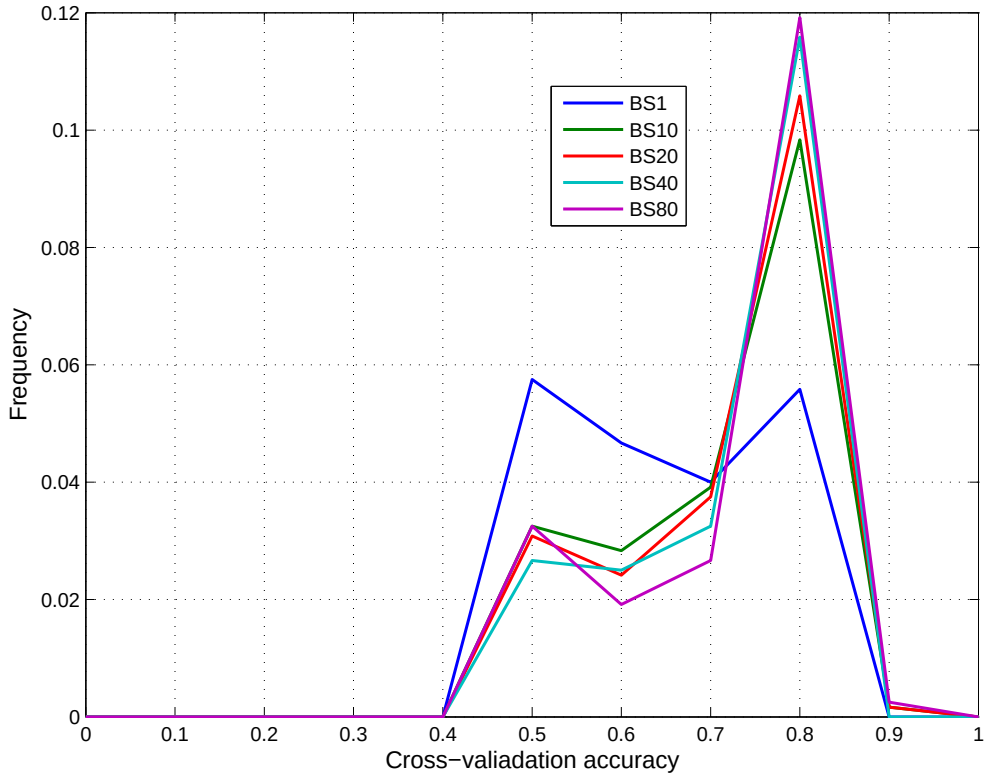


Figure 4.2: Distribution of CV accuracy at different levels of bootstrap sampling

Figure 4.3 shows the histogram of the frequency of each feature from the linear SVM. The top 10 most frequent features were V_{69}, V_{70}, V_{71}, V_{72}, V_{48}, V_{49}, V_{50}, V_{51}, and V_{53}.

4.3 Discussion

Despite considering each feature to add independently of the others, the forward Spearman-based feature selector performed better than the backwards RFE feature selector. This suggests that the univariate addition of features and the assumption of independent features is not a large detriment to the performance of the classifier.

The small sample size of the data could have led to reduced classifier performance. We modified the original dataset by treating primary and nodal targets the same. This was done in order to increase the amount of usable data. In the original data set, each patient could have a primary target, nodal target, or both. Each target has its own response label associated with it. Due to the small amount of sample data, we were forced to treat the primary target and nodal targets as examples of the underlying source data. In reality, primary targets and nodal targets behave differently. This confounding effect could have contributed to the degradation of classifier performance. Also, the generality of our results depends on if the distribution of response of the true patient population is similar to this sample of patients. LOOCV is also known to possess high variance in its estimated performance (Efron and Tibshirani, 1997), therefore a method with a better estimated performance may not actually perform better in reality.

When the linear SVM classifier was presented with more features from which to construct a feature subset, the overall performance seems to degrade. This could be due to overfitting on the training data

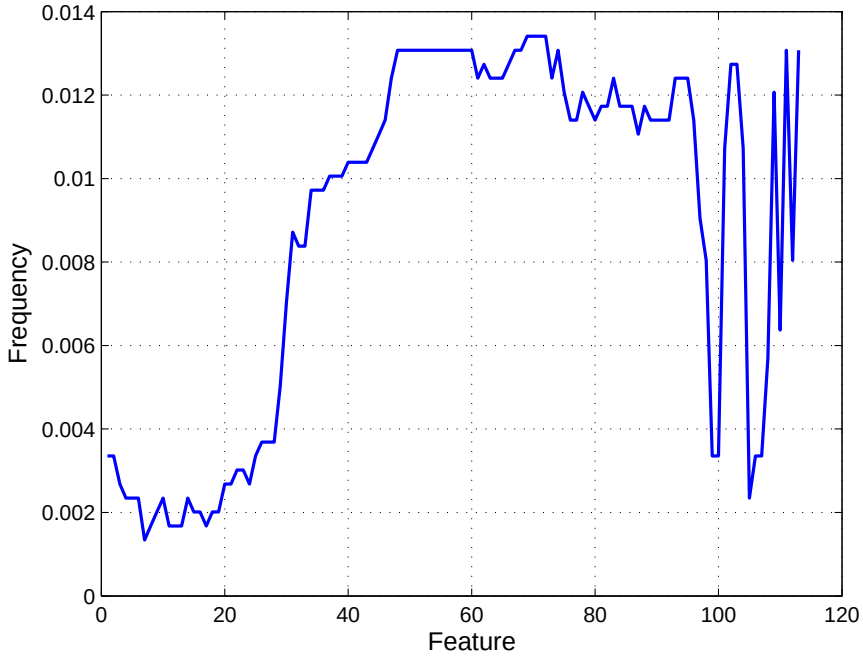


Figure 4.3: Frequency of each feature in Scan 1 for linear SVM

and resulting in misclassification. A larger dataset may allow for better use of additional features and their inherent relationships.

The list of most common features suggests that the IVH holds prognostic value in predicting patient response. However, it appears that within these high frequency features are sequences of highly related features. This could just be a result of the way the IVH has been discretized. If we decided use a coarser discretization of the IVH, for example at every 10^{th} value only, we would obtain another set of most frequent features.

Chapter 5

Conclusions and future work

In this thesis, we investigated methods of incorporating FDG-PET data into the treatment planning process. We considered its use in the fluence map optimization problem to design biologically-guided treatment plans, and also its prognostic value in predicting patient response to treatment.

The field of radiation oncology is moving towards a more patient-specific era of treatment. With the advent of IMRT, clinicians were able to design treatment plans that sculpted the dose to each patient's unique anatomy. And now with the use of PET, clinicians have yet another way to tailor their treatments for each individual patient. Combining the prognostic capability of PET with biologically-guided IMRT could introduce promising ways to adapt to the patient's changing functional activity throughout the course of treatment. For instance, it would be the ideal situation if we were able to predict the progression of a tumor at any stage of the treatment process based on PET imaging data. We then use this biological data to inform the optimal prescription dose and then try to deliver that dose and then update our beliefs on the progression of the tumor for the next time step. This biologically adaptive treatment planning could allow the treatments to be even more customized for patients since indicators of tumor biology are taken into account.

Robust biologically-guided intensity-modulated radiation therapy

In this thesis, we developed a robust optimization framework for biologically-guided IMRT. We demonstrate its effectiveness in the face of uncertainty with respect to the SUV observed in the original PET image which then translates to uncertainty in the prescription dose. We investigated abstract mathematical realizations of uncertainty and also a more realistic uncertainty scenario based on the partial volume effect. In all of our simulations, the robust model outperformed the nominal model in both objective function value and TCP value. Even if we adjust our assumed parameter values, the robust model was still able to perform within reasonable limits, while the nominal model's performance degraded rapidly.

For future work, this robust framework should be tested on more clinical patient data to further validate the approach. Utilization of more realistic SUV to dose functions would improve the fidelity of the work. We also did not consider sensitivity to the conversion function from oxygen to radiosensitivity parameters.

This work did not consider the effects of intra-fraction or inter-fraction motion on the delivery of a

highly heterogeneous dose. Successful delivery of a plan with potentially many sharp dose gradients in different locations of the tumor relies on accurate targeting of the tumor.

Other tracers could provide more realistic spatial distributions of radioresistance. Our provided dataset used FDG as its tracer and we assumed a relationship between FDG uptake and hypoxia to obtain a map of radioresistance. However, other tracers such fluoromisonidazole (FMISO) and fluoroazomycinarabinofuranoside (FAZA) are much more suitable for measuring hypoxia (Bentzen and Grégoire, 2011). Since our formulations optimize over deviations from prescription dose, switching to other tracers or even other functional imaging modalities such as magnetic resonance imaging (MRI) or single photon emission computed tomography (SPECT) would be simple. The only thing that would change would be the conversion functions from image values to prescription dose.

All of this robust work was performed assuming a single planning instance at the start of treatment. An promising direction for this work would be to incorporate the robust biological optimization with an adaptive framework. One expects the SUV to change from scan to scan. By adapting to each scan, it is hoped that we could tailor the treatment even better for the patient and potentially reduce the requirement for higher doses as the treatment progresses and the patient condition improves. The fusion of robust biological treatment planning and adaptive treatment planning would be an important step closer to the goal of highly customized treatment planing for the individual.

Lung cancer response classification using pattern recognition

In this thesis, we applied machine learning algorithms to 4D gated PET-FDG/CT scans. We identified classifiers which could outperform a naive classifier. We were able to perform well despite only using the pre-treatment scan. In fact, it appears that too many scans may deteriorate the performance of the classifier, possibly due to overfitting. Our work shows the potential of using machine learning algorithms to predict patient response. This could hopefully lead to more adaptive treatment plans, where the clinician would adapt the treatment based on the prediction provided at time t in the treatment.

For future work, collection of more data is paramount. Typically in machine learning, more data beats a more complicated algorithm. With more data, we may be able to reduce the potential confounding effect of grouping primary tumors and nodal tumors together.

Appendix A

Complete cross-validation accuracy tables

The tables contained in this Appendix detail all the different combinations of hyperparameter settings used in the patient response prediction experiments. Each table represents a given classifer and set of scans. The decimal values are the resultant CV accuracy values. The column "Majority/Minority Ratio" indicates if oversampling was performed, and if so to what degree. The original majority to minority class ratio was $\frac{36}{5} = 7.2$. For example, a ratio value of 1 indicates that an additional 31 minority class examples were replicated from the original 5 and added to the dataset.

		Feature Selection		
Majority/Minority Ratio	Num bootstrap	PCA	RFE	SPEAR
Original	1	0.7317	0.7805	0.6829
	10	0.878	0.878	0.8537
	20	0.878	0.878	0.8537
	40	0.878	0.878	0.8537
	80	0.878	0.878	0.8537
4	1	0.561	0.6341	0.5854
	10	0.6098	0.8049	0.7073
	20	0.6098	0.8537	0.7073
	40	0.561	0.8293	0.7561
	80	0.5366	0.8293	0.7317
3	1	0.6341	0.7317	0.561
	10	0.6341	0.7317	0.6585
	20	0.6829	0.7805	0.5366
	40	0.6341	0.878	0.6585
	80	0.5122	0.8537	0.5854
2	1	0.6098	0.6098	0.561
	10	0.561	0.6341	0.5366
	20	0.561	0.7317	0.561
	40	0.5366	0.7073	0.5122
	80	0.5854	0.7561	0.5122
1	1	0.5366	0.5366	0.5122
	10	0.6098	0.6098	0.5366
	20	0.561	0.5122	0.7073
	40	0.6341	0.6585	0.5854
	80	0.5366	0.5366	0.5854

Table A.1: L2-regularized logistic regression, Scan 1

		Feature Selection		
Majority/Minority ratio	Num bootstrap	PCA	RFE	SPEAR
Original	1	0.7317	0.7805	0.7317
	10	0.878	0.878	0.878
	20	0.878	0.878	0.878
	40	0.878	0.878	0.878
	80	0.878	0.878	0.878
4	1	0.5366	0.5366	0.7073
	10	0.561	0.8537	0.7805
	20	0.6585	0.7805	0.8293
	40	0.6585	0.8293	0.878
	80	0.6585	0.8049	0.878
3	1	0.6098	0.6829	0.6098
	10	0.5366	0.6585	0.7561
	20	0.561	0.6585	0.7805
	40	0.5366	0.7805	0.7561
	80	0.561	0.8293	0.8293
2	1	0.5366	0.5122	0.5366
	10	0.6829	0.5366	0.6585
	20	0.5122	0.561	0.5366
	40	0.6585	0.6585	0.5366
	80	0.7073	0.6098	0.5122
1	1	0.5122	0.5122	0.5854
	10	0.5122	0.5854	0.5366
	20	0.6098	0.5854	0.6829
	40	0.5366	0.6829	0.5366
	80	0.5854	0.5854	0.5366

Table A.2: L1-regularized logistic regression, Scan 1

		Feature Selection		
Majority/Minority ratio	Num bootstrap	PCA	RFE	SPEAR
Original	1	0.878	0.8293	0.7317
	10	0.878	0.878	0.7805
	20	0.878	0.8537	0.7805
	40	0.878	0.8537	0.7805
	80	0.878	0.8537	0.8049
4	1	0.8049	0.5854	0.7805
	10	0.8537	0.6585	0.8049
	20	0.878	0.6585	0.8049
	40	0.878	0.7073	0.8537
	80	0.878	0.7073	0.8293
3	1	0.8293	0.561	0.8537
	10	0.878	0.7073	0.9268
	20	0.8537	0.6829	0.8537
	40	0.878	0.7073	0.8293
	80	0.8537	0.6585	0.8537
2	1	0.8537	0.6585	0.878
	10	0.878	0.7561	0.9024
	20	0.878	0.7805	0.878
	40	0.878	0.8293	0.8537
	80	0.878	0.8049	0.878
1	1	0.8293	0.6829	0.8293
	10	0.8537	0.7805	0.8537
	20	0.878	0.878	0.8537
	40	0.878	0.878	0.8537
	80	0.878	0.878	0.878

Table A.3: Linear SVM, Scan 1

		Feature Selection		
Majority/Minority ratio	Num bootstrap	PCA	RFE	SPEAR
Original	1	0.878	0.878	0.8049
	10	0.878	0.878	0.878
	20	0.878	0.878	0.878
	40	0.878	0.878	0.878
	80	0.878	0.878	0.878
4	1	0.8293	0.878	0.6585
	10	0.878	0.878	0.8293
	20	0.878	0.878	0.878
	40	0.878	0.878	0.878
	80	0.878	0.878	0.878
3	1	0.878	0.8537	0.6098
	10	0.878	0.878	0.8537
	20	0.878	0.878	0.878
	40	0.878	0.878	0.7073
	80	0.878	0.878	0.6098
2	1	0.878	0.878	0.7561
	10	0.878	0.878	0.7805
	20	0.878	0.878	0.8537
	40	0.878	0.878	0.8537
	80	0.878	0.878	0.878
1	1	0.878	0.878	0.7317
	10	0.878	0.878	0.6098
	20	0.878	0.878	0.8293
	40	0.878	0.878	0.7805
	80	0.878	0.878	0.8293

Table A.4: Radial basis kernel SVM, Scan 1

		Feature Selection		
Majority/Minority ratio	Num bootstrap	PCA	RFE	SPEAR
Original	1	0.7805	0.7317	0.7317
	10	0.878	0.878	0.878
	20	0.878	0.878	0.878
	40	0.878	0.878	0.878
	80	0.878	0.878	0.878
4	1	0.6098	0.6585	0.6341
	10	0.5854	0.8293	0.6829
	20	0.7317	0.8537	0.6341
	40	0.6341	0.7805	0.6829
	80	0.5122	0.8537	0.7317
3	1	0.5366	0.5854	0.561
	10	0.6585	0.7073	0.5366
	20	0.7317	0.7073	0.5122
	40	0.7317	0.8293	0.561
	80	0.6098	0.8293	0.6829
2	1	0.6098	0.5854	0.5122
	10	0.7561	0.7561	0.6341
	20	0.8537	0.7561	0.5122
	40	0.878	0.8049	0.5122
	80	0.8049	0.7073	0.5854
1	1	0.5366	0.5122	0.6098
	10	0.5366	0.6829	0.561
	20	0.5854	0.561	0.561
	40	0.6098	0.5122	0.561
	80	0.5122	0.5854	0.5854

Table A.5: L2-regularized logistic regression, Scan 1-2

		Feature Selection		
Majority/Minority ratio	Num bootstrap	PCA	RFE	SPEAR
Original	1	0.7073	0.7317	0.7317
	10	0.8293	0.878	0.878
	20	0.7561	0.878	0.878
	40	0.8293	0.878	0.878
	80	0.7805	0.878	0.878
4	1	0.5366	0.6341	0.7317
	10	0.561	0.8049	0.8537
	20	0.6341	0.8293	0.7805
	40	0.6829	0.8293	0.878
	80	0.5122	0.8293	0.8537
3	1	0.5854	0.561	0.561
	10	0.5122	0.6585	0.6829
	20	0.6098	0.7561	0.7317
	40	0.6341	0.8537	0.8537
	80	0.5122	0.7805	0.8293
2	1	0.7317	0.5366	0.5854
	10	0.7317	0.7561	0.5122
	20	0.7317	0.6829	0.5854
	40	0.8293	0.7805	0.6098
	80	0.7805	0.6585	0.5854
1	1	0.5366	0.5854	0.5854
	10	0.5366	0.7561	0.5122
	20	0.5122	0.5366	0.5122
	40	0.5122	0.5366	0.5122
	80	0.6098	0.5122	0.5366

Table A.6: L1-regularized logistic regression, Scan 1-2

		Feature Selection		
Majority/Minority ratio	Num bootstrap	PCA	RFE	SPEAR
Original	1	0.878	0.7073	0.7805
	10	0.878	0.7073	0.8293
	20	0.878	0.8293	0.8049
	40	0.878	0.7805	0.8537
	80	0.878	0.9024	0.8537
4	1	0.8537	0.6585	0.878
	10	0.878	0.7073	0.8049
	20	0.878	0.7073	0.9024
	40	0.878	0.5854	0.8537
	80	0.878	0.6098	0.8537
3	1	0.7317	0.5366	0.7317
	10	0.878	0.7073	0.878
	20	0.878	0.6585	0.878
	40	0.878	0.6585	0.8049
	80	0.878	0.6585	0.8049
2	1	0.7805	0.5122	0.6341
	10	0.878	0.6829	0.8537
	20	0.878	0.7561	0.8293
	40	0.878	0.7317	0.8537
	80	0.8537	0.7561	0.9024
1	1	0.7317	0.6585	0.7317
	10	0.8537	0.8537	0.8537
	20	0.8537	0.8293	0.8537
	40	0.8537	0.8537	0.878
	80	0.878	0.8537	0.8537

Table A.7: Linear SVM, Scan 1-2

		Feature Selection		
Majority/Minority ratio	Num bootstrap	PCA	RFE	SPEAR
Original	1	0.878	0.878	0.7561
	10	0.878	0.878	0.8293
	20	0.878	0.878	0.8293
	40	0.878	0.878	0.8537
	80	0.878	0.878	0.8537
4	1	0.878	0.878	0.878
	10	0.878	0.878	0.878
	20	0.878	0.878	0.561
	40	0.878	0.878	0.8537
	80	0.878	0.878	0.878
3	1	0.878	0.878	0.878
	10	0.878	0.878	0.878
	20	0.878	0.878	0.8293
	40	0.878	0.878	0.8049
	80	0.878	0.878	0.9024
2	1	0.878	0.8537	0.8049
	10	0.878	0.878	0.8537
	20	0.878	0.878	0.7317
	40	0.878	0.878	0.8537
	80	0.878	0.878	0.8049
1	1	0.878	0.8537	0.5122
	10	0.878	0.878	0.5366
	20	0.878	0.878	0.6585
	40	0.878	0.878	0.6341
	80	0.878	0.878	0.7561

Table A.8: Radial basis kernel SVM, Scan 1-2

		Feature Selection		
Majority/Minority ratio	Num bootstrap	PCA	RFE	SPEAR
Original	1	0.8049	0.7317	0.8049
	10	0.878	0.878	0.878
	20	0.878	0.878	0.878
	40	0.878	0.878	0.878
	80	0.878	0.878	0.878
4	1	0.6098	0.6341	0.5366
	10	0.6341	0.8293	0.7805
	20	0.5366	0.8537	0.7073
	40	0.5854	0.878	0.7561
	80	0.6341	0.8293	0.7805
3	1	0.6098	0.5366	0.6829
	10	0.6098	0.8293	0.6341
	20	0.5122	0.8049	0.7317
	40	0.5366	0.8293	0.6098
	80	0.5366	0.8537	0.6098
2	1	0.5366	0.5854	0.5854
	10	0.6585	0.7805	0.6098
	20	0.6098	0.7317	0.5366
	40	0.5366	0.7317	0.561
	80	0.5122	0.7317	0.5122
1	1	0.5122	0.5122	0.5366
	10	0.6098	0.5122	0.5122
	20	0.5122	0.5366	0.5122
	40	0.6585	0.5366	0.5122
	80	0.5122	0.5122	0.6585

Table A.9: L2-regularized logistic regression, Scan 1-3

		Feature Selection		
Majority/Minority ratio	Num bootstrap	PCA	RFE	SPEAR
Original	1	0.8049	0.7561	0.8049
	10	0.878	0.8537	0.878
	20	0.878	0.8537	0.878
	40	0.878	0.8537	0.878
	80	0.878	0.8537	0.878
4	1	0.5366	0.6341	0.6829
	10	0.7805	0.8049	0.7805
	20	0.6585	0.7805	0.8537
	40	0.6829	0.8537	0.878
	80	0.6585	0.8049	0.8293
3	1	0.6829	0.6098	0.6341
	10	0.7317	0.7805	0.7805
	20	0.6341	0.7561	0.6829
	40	0.7073	0.8049	0.7805
	80	0.5854	0.8537	0.8293
2	1	0.5122	0.7073	0.5366
	10	0.561	0.6585	0.5122
	20	0.5366	0.7073	0.6098
	40	0.6829	0.7317	0.5854
	80	0.5122	0.7073	0.6341
1	1	0.5122	0.5122	0.561
	10	0.5366	0.5854	0.561
	20	0.5122	0.5366	0.7073
	40	0.5366	0.5366	0.5854
	80	0.5122	0.5366	0.5122

Table A.10: L1-regularized logistic regression, Scan 1-3

		Feature Selection		
Majority/Minority ratio	Num bootstrap	PCA	RFE	SPEAR
Original	1	0.878	0.7073	0.8537
	10	0.878	0.7805	0.8537
	20	0.878	0.7073	0.8293
	40	0.878	0.7317	0.8049
	80	0.878	0.7561	0.8537
4	1	0.8049	0.5122	0.6098
	10	0.7561	0.7805	0.7073
	20	0.8293	0.7073	0.7561
	40	0.8049	0.7561	0.8293
	80	0.7561	0.7805	0.7317
3	1	0.7561	0.6829	0.5122
	10	0.7805	0.5854	0.7317
	20	0.8049	0.6341	0.7073
	40	0.8537	0.7561	0.7805
	80	0.7805	0.8049	0.7317
2	1	0.7073	0.6098	0.6829
	10	0.8293	0.7073	0.7073
	20	0.8049	0.7805	0.7561
	40	0.8293	0.6829	0.7805
	80	0.8293	0.8537	0.7317
1	1	0.7561	0.6341	0.5854
	10	0.878	0.7805	0.8537
	20	0.878	0.8049	0.8537
	40	0.878	0.8537	0.7805
	80	0.878	0.878	0.878

Table A.11: Linear SVM, Scan 1-3

		Feature Selection		
Majority/Minority ratio	Num bootstrap	PCA	RFE	SPEAR
Original	1	0.878	0.878	0.7561
	10	0.878	0.878	0.8049
	20	0.878	0.878	0.8537
	40	0.878	0.878	0.878
	80	0.878	0.878	0.8537
4	1	0.878	0.878	0.6341
	10	0.878	0.878	0.7561
	20	0.878	0.878	0.6098
	40	0.878	0.878	0.6341
	80	0.878	0.878	0.878
3	1	0.878	0.878	0.8049
	10	0.878	0.878	0.7317
	20	0.878	0.878	0.7561
	40	0.878	0.878	0.878
	80	0.878	0.878	0.878
2	1	0.878	0.878	0.7073
	10	0.878	0.878	0.7317
	20	0.878	0.878	0.7317
	40	0.878	0.878	0.878
	80	0.878	0.878	0.8537
1	1	0.878	0.878	0.561
	10	0.878	0.878	0.6341
	20	0.878	0.878	0.561
	40	0.878	0.878	0.6829
	80	0.878	0.878	0.7073

Table A.12: Radial basis kernel SVM, Scan 1-3

		Feature Selection		
Majority/Minority ratio	Num bootstrap	PCA	RFE	SPEAR
Original	1	0.7805	0.6829	0.7805
	10	0.878	0.878	0.878
	20	0.878	0.878	0.878
	40	0.878	0.878	0.878
	80	0.878	0.878	0.878
4	1	0.5854	0.6585	0.6341
	10	0.6585	0.8293	0.7073
	20	0.6829	0.8293	0.7317
	40	0.6829	0.8537	0.7317
	80	0.6341	0.8537	0.6585
3	1	0.561	0.7073	0.6341
	10	0.5854	0.8293	0.5854
	20	0.6098	0.8537	0.6829
	40	0.6341	0.8293	0.6098
	80	0.6341	0.8537	0.6341
2	1	0.5122	0.6341	0.5366
	10	0.6585	0.6585	0.5854
	20	0.7317	0.7317	0.561
	40	0.7561	0.7317	0.7073
	80	0.7073	0.8049	0.5854
1	1	0.561	0.5366	0.5366
	10	0.561	0.6098	0.5854
	20	0.5854	0.561	0.6098
	40	0.5854	0.5122	0.6585
	80	0.6098	0.6098	0.6341

Table A.13: L2-regularized logistic regression, Scan 1-4

		Feature Selection		
Majority/Minority ratio	Num bootstrap	PCA	RFE	SPEAR
Original	1	0.7805	0.6341	0.7805
	10	0.878	0.8537	0.878
	20	0.878	0.878	0.878
	40	0.878	0.878	0.878
	80	0.878	0.878	0.878
4	1	0.5366	0.6098	0.7561
	10	0.5854	0.8049	0.7561
	20	0.561	0.8537	0.8293
	40	0.6829	0.8293	0.8293
	80	0.5122	0.878	0.8537
3	1	0.6341	0.6829	0.7317
	10	0.5854	0.7805	0.7317
	20	0.6341	0.878	0.7805
	40	0.6098	0.8293	0.7805
	80	0.5122	0.8049	0.8049
2	1	0.5854	0.6098	0.5366
	10	0.561	0.6341	0.5854
	20	0.6585	0.7561	0.5366
	40	0.6585	0.7561	0.5122
	80	0.7073	0.7561	0.561
1	1	0.5366	0.5122	0.561
	10	0.561	0.5854	0.6829
	20	0.6829	0.5854	0.561
	40	0.5122	0.5122	0.5854
	80	0.5366	0.6585	0.5854

Table A.14: L1-regularized logistic regression, Scan 1-4

		Feature Selection		
Majority/Minority ratio	Num bootstrap	PCA	RFE	SPEAR
Original	1	0.878	0.6098	0.8049
	10	0.878	0.6585	0.8537
	20	0.878	0.7317	0.8049
	40	0.878	0.7073	0.8049
	80	0.878	0.7805	0.7805
4	1	0.8049	0.6341	0.6098
	10	0.8537	0.8049	0.7073
	20	0.8537	0.9024	0.8293
	40	0.8537	0.8049	0.7317
	80	0.8537	0.878	0.8293
3	1	0.8537	0.6341	0.6829
	10	0.8537	0.7317	0.8293
	20	0.8537	0.8293	0.7805
	40	0.878	0.7805	0.8293
	80	0.878	0.7561	0.7561
2	1	0.7561	0.5366	0.6829
	10	0.878	0.7561	0.8293
	20	0.8537	0.7805	0.8293
	40	0.878	0.7805	0.7805
	80	0.8537	0.8049	0.7805
1	1	0.7317	0.6585	0.5366
	10	0.878	0.7805	0.7561
	20	0.8537	0.878	0.8293
	40	0.878	0.8537	0.8293
	80	0.878	0.8537	0.878

Table A.15: Linear SVM, Scan 1-4

		Feature Selection		
Majority/Minority ratio	Num bootstrap	PCA	RFE	SPEAR
Original	1	0.878	0.8049	0.7805
	10	0.878	0.8293	0.8049
	20	0.878	0.878	0.878
	40	0.878	0.8293	0.8537
	80	0.878	0.8293	0.878
4	1	0.878	0.7561	0.6585
	10	0.878	0.8537	0.878
	20	0.878	0.878	0.7073
	40	0.878	0.878	0.7561
	80	0.878	0.878	0.878
3	1	0.878	0.878	0.6341
	10	0.878	0.878	0.6585
	20	0.878	0.878	0.7805
	40	0.878	0.878	0.7073
	80	0.878	0.878	0.8049
2	1	0.878	0.878	0.7561
	10	0.878	0.878	0.7073
	20	0.878	0.878	0.8537
	40	0.878	0.878	0.7073
	80	0.878	0.878	0.7805
1	1	0.878	0.878	0.6341
	10	0.878	0.878	0.5366
	20	0.878	0.878	0.6829
	40	0.878	0.878	0.8049
	80	0.878	0.878	0.7561

Table A.16: Radial basis kernel SVM, Scan 1-4

Bibliography

Hugo JWL Aerts, Angela AW van Baardwijk, Steven F Petit, Claudia Offermann, Judith van Loon, Ruud Houben, Anne-Marie C Dingemans, Rinus Wanders, Liesbeth Boersma, Jacques Borger, et al. Identification of residual metabolic-active areas within individual nsclc tumours using a pre-radiotherapy¡ sup¿ 18¡/sup¿ fluorodeoxyglucose-pet-ct scan. *Radiotherapy and Oncology*, 91(3):386–392, 2009.

R. Airley, J. Loncaster, S. Davidson, M. Bromley, S. Roberts, A. Patterson, R. Hunter, I. Stratford, and C. West. Glucose transporter glut-1 expression correlates with tumor hypoxia and predicts metastasis-free survival in advanced carcinoma of the cervix. *Clinical cancer research*, 7(4):928–934, 2001.

M Alber, F Paulsen, SM Eschmann, and HJ Machulla. On biologically conformal boost dose optimization. *Physics in medicine and biology*, 48(2):N31, 2003.

Songjoon Baek, Chen-An Tsai, and James J Chen. Development of biomarker classifiers from high-dimensional data. *Briefings in bioinformatics*, 10(5):537–546, 2009.

Simona Ben-Haim and Peter Ell. 18f-fdg pet and pet/ct in the evaluation of cancer treatment response. *Journal of Nuclear Medicine*, 50(1):88–99, 2009.

Søren M Bentzen and Vincent Grégoire. Molecular imaging–based dose painting: A novel paradigm for radiation therapy prescription. In *Seminars in radiation oncology*, volume 21, pages 101–110. Elsevier, 2011.

Abi Berger. How does it work?: Positron emission tomography. *BMJ: British Medical Journal*, 326 (7404):1449, 2003.

Thierry Berghmans, Michèle Dusart, Marianne Paesmans, Claude Hossein-Foucher, Irene Buvat, Catherine Castaigne, Arnaud Scherpereel, Céline Mascaux, Michel Moreau, Martine Roelandts, et al. Primary tumor standardized uptake value (suvmax) measured on fluorodeoxyglucose positron emission tomography (fdg-pet) is of prognostic value for survival in non-small cell lung cancer (nsclc): a systematic review and meta-analysis (ma) by the european lung cancer working party for the iaslc lung cancer staging project. *Journal of Thoracic Oncology*, 3(1):6–12, 2008.

Dimitris Bertsimas and Melvyn Sim. The price of robustness. *Operations research*, 52(1):35–53, 2004.

Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*, volume 1. springer New York, 2006.

Ronald Boellaard. Standards for pet image acquisition and quantitative data analysis. *Journal of nuclear medicine*, 50(Suppl 1):11S–20S, 2009.

Thomas Bortfeld, Timothy CY Chan, Alexei Trofimov, and John N Tsitsiklis. Robust management of motion uncertainty in intensity-modulated radiation therapy. *Operations Research*, 56(6):1461–1473, 2008.

Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152. ACM, 1992.

Stephen R Bowen, Matthew J Nyflot, Michael Gensheimer, Kristi RG Hendrickson, Paul E Kinahan, George A Sandison, and Shilpen A Patel. Challenges and opportunities in patient-specific, motion-managed and pet/ct-guided radiation therapy of lung cancer: review and perspective. *Clinical and translational medicine*, 1(1):1–17, 2012.

A. Brahme and AK Argren. Optimal dose distribution for eradication of heterogeneous tumors. *Acta Oncologica*, 26(5):377–385, 1987.

Anders Brahme. Optimized radiation therapy based on radiobiological objectives. In *Seminars in radiation oncology*, volume 9, pages 35–47. Elsevier, 1999.

Leo Breiman. Bagging predictors. *Machine learning*, 24(2):123–140, 1996.

Timothy CY Chan, Thomas Bortfeld, and John N Tsitsiklis. A robust approach to imrt optimization. *Physics in medicine and biology*, 51(10):2567, 2006.

Chih-Chung Chang and Chih-Jen Lin. Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27, 2011.

Wei Chen, Jan Unkelbach, Alexei Trofimov, Thomas Madden, Hanne Kooy, Thomas Bortfeld, and David Craft. Including robustness in multi-criteria optimization for intensity-modulated proton therapy. *Physics in Medicine and Biology*, 57(3):591, 2012.

Nicolas Christian, John A Lee, Anne Bol, Marc De Bast, Bénédicte Jordan, and Vincent Grégoire. The limitation of pet imaging for biological adaptive-imrt assessed in animal models. *Radiotherapy and Oncology*, 91(1):101–106, 2009.

Millie Chu, Yuriy Zinchenko, Shane G Henderson, and Michael B Sharpe. Robust optimization for intensity modulated radiation therapy treatment planning under uncertainty. *Physics in Medicine and Biology*, 50(23):5463, 2005.

Gregory W Corder and Dale I Foreman. *Nonparametric statistics for non-statisticians: a step-by-step approach*. Wiley. com, 2009.

Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

Joseph A Cruz and David S Wishart. Applications of machine learning in cancer prediction and prognosis. *Cancer informatics*, 2:59, 2006.

SK Das, MM Miften, S Zhou, M Bell, MT Munley, CS Whiddon, O Craciunescu, AH Baydush, T Wong, JG Rosenman, et al. Feasibility of optimizing the dose distribution in lung tumors using fluorine-18-fluorodeoxyglucose positron emission tomography and single photon emission computed tomography guided dose prescriptions. *Medical physics*, 31:1452, 2004.

Joseph O Deasy, Angel I Blanco, and Vanessa H Clark. Cerr: a computational environment for radiotherapy research. *Medical physics*, 30:979, 2003.

Christophe Dooms, Angela van Baardwijk, Eric Verbeken, Robert Jan van Suylen, Sigrid Stroobants, Dirk De Ruysscher, and Johan Vansteenkiste. Association between 18f-fluoro-2-deoxy-d-glucose uptake values and tumor vitality: prognostic value of positron emission tomography in early-stage non-small cell lung cancer. *Journal of Thoracic Oncology*, 4(7):822–828, 2009.

Robert J Downey, Timothy Akhurst, Mithat Gonen, Alain Vincent, Manjit S Bains, Steven Larson, and Valerie Rusch. Preoperative f-18 fluorodeoxyglucose-positron emission tomography maximal standardized uptake value predicts survival after lung cancer resection. *Journal of Clinical Oncology*, 22(16): 3255–3260, 2004.

Bradley Efron and Robert Tibshirani. Improvements on cross-validation: the 632+ bootstrap method. *Journal of the American Statistical Association*, 92(438):548–560, 1997.

Matthias Ehrgott, Çiğdem Güler, Horst W Hamacher, and Lizhen Shao. Mathematical optimization in intensity modulated radiation therapy. *4OR: a Quarterly Journal of Operations Research*, 6(3): 199–262, 2008.

I El Naqa, PW Grigsby, A Apte, E Kidd, E Donnelly, D Khullar, S Chaudhari, D Yang, M Schmitt, Richard Laforest, et al. Exploring feature-based approaches in pet images for predicting cancer treatment outcomes. *Pattern recognition*, 42(6):1162–1171, 2009.

SY El Sharouni, HB Kal, and JJ Battermann. Tumour control probability of stage iii inoperable non-small cell lung tumours after sequential chemo-radiotherapy. *Anticancer research*, 25(6C):4655–4661, 2005.

SM Eschmann, G Friedel, F Paulsen, M Reimold, T Hehr, W Budach, J Scheiderbauer, HJ Machulla, H Dittmann, R Vonthein, et al. Is standardised 18f-fdg uptake value an outcome predictor in patients with stage iii non-small cell lung cancer? *European journal of nuclear medicine and molecular imaging*, 33(3):263–269, 2006.

R.E. Fan, K.W. Chang, C.J. Hsieh, X.R. Wang, and C.J. Lin. LIBLINEAR: A library for large linear classification. *The Journal of Machine Learning Research*, 9:1871–1874, 2008.

J.F. Fowler. The linear-quadratic formula and progress in fractionated radiotherapy. *British Journal of Radiology*, 62(740):679–694, 1989.

Albin Fredriksson, Anders Forsgren, and Björn Hårdemark. Minimax optimization for handling range and setup uncertainties in proton therapy. *Medical Physics*, 38:1672, 2011.

Sanjiv Sam Gambhir. Molecular imaging of cancer with positron emission tomography. *Nature Reviews Cancer*, 2(9):683–693, 2002.

Boone Goodgame, Giancarlo A Pillot, Zhiyun Yang, Jabi Shriki, Bryan F Meyers, Jennifer Zoole, Feng Gao, Farrokh Dehdashti, Alexander Patterson, Barry A Siegel, et al. Prognostic value of preoperative positron emission tomography in resected stage i non-small cell lung cancer. *Journal of Thoracic Oncology*, 3(2):130–134, 2008.

P.W. Grigsby, B.A. Siegel, F. Dehdashti, J. Rader, and I. Zoberi. Posttherapy [18f] fluorodeoxyglucose positron emission tomography in carcinoma of the cervix: response and outcome. *Journal of clinical oncology*, 22(11):2167–2171, 2004.

Isabelle Guyon. *Feature extraction: foundations and applications*, volume 207. Springer, 2006.

Dirk Hellwig, Thomas P Graeter, Dieter Ukena, Andreas Groeschel, Gerhard W Sybrecht, Hans-Joachim Schaefers, and Carl-Martin Kirsch. 18f-fdg pet for mediastinal staging of lung cancer: which suv threshold makes sense? *Journal of Nuclear Medicine*, 48(11):1761–1766, 2007.

Iñaki Inza, Borja Calvo, Rubén Armañanzas, Endika Bengoetxea, Pedro Larrañaga, and José A Lozano. Machine learning: an indispensable tool in bioinformatics. In *Bioinformatics methods in clinical research*, pages 25–48. Springer, 2010.

Ahmedin Jemal, Freddie Bray, Melissa M Center, Jacques Ferlay, Elizabeth Ward, and David Forman. Global cancer statistics. *CA: a cancer journal for clinicians*, 61(2):69–90, 2011.

IT Jolliffe. Principal component analysis. 1986. *Spring-verlag, New York*, 1986.

Malik E Juweid and Bruce D Cheson. Positron-emission tomography and assessment of cancer therapy. *New England Journal of Medicine*, 354(5):496–507, 2006.

G. Kåver, B.K. Lind, J. Löf, A. Liander, and A. Brahme. Stochastic optimization of intensity modulated radiotherapy to account for uncertainties in patient sensitivity. *Physics in medicine and biology*, 44 (12):2955, 1999.

Minsun Kim, Archis Ghate, and Mark H Phillips. A stochastic control formalism for dynamic biologically conformal radiation therapy. *European Journal of Operational Research*, 219(3):541–556, 2012.

Yusung Kim and Wolfgang A Tomé. Risk-adaptive optimization: selective boosting of high-risk tumor subvolumes. *International Journal of Radiation Oncology* Biology* Physics*, 66(5):1528–1542, 2006.

D Levin-Plotnik and RJ Hamilton. Optimization of tumour control probability for heterogeneous tumours in fractionated radiotherapy treatment protocols. *Physics in medicine and biology*, 49(3):407, 2004.

X Allen Li, Markus Alber, Joseph O Deasy, Andrew Jackson, Kyung-Wook Ken Jee, Lawrence B Marks, Mary K Martel, Charles Mayo, Vitali Moiseenko, Alan E Nahum, et al. The use and qa of biologically related models for treatment planning: Short report of the tg-166 of the therapy physics committee of the aapm. *Medical Physics*, 39:1386, 2012.

Gerald Lim, Andrea Bezjak, Jane Higgins, Doug Moseley, Andrew J Hope, Alex Sun, John BC Cho, Anthony M Brade, Clement Ma, and Jean-Pierre Bissonnette. Tumor regression and positional changes in non-small cell lung cancer during radical radiotherapy. *Journal of Thoracic Oncology*, 6(3):531–536, 2011.

C.C. Ling, J. Humm, S. Larson, H. Amols, Z. Fuks, S. Leibel, and J.A. Koutcher. Towards multidimensional radiotherapy (md-crt): biological imaging and biological conformality. *International Journal of Radiation Oncology* Biology* Physics*, 47(3):551–560, 2000.

Wei Liu, Xiaodong Zhang, Yupeng Li, and Radhe Mohan. Robust optimization of intensity modulated proton therapy. *Medical physics*, 39:1079, 2012.

Michael MacManus, Ursula Nestle, Kenneth E Rosenzweig, Ignasi Carrio, Cristina Messa, Otakar Belohlavek, Massimo Danna, Tomio Inoue, Elizabeth Deniaud-Alexandre, Stefano Schipani, et al. Use of pet and pet/ct for radiation therapy planning: Iaea expert report 2006–2007. *Radiotherapy and Oncology*, 91(1):85–94, 2009.

George D Magoulas and Andriana Prentza. Machine learning in medical applications. In *Machine Learning and its applications*, pages 300–307. Springer, 2001.

Lawrence B Marks, Ellen D Yorke, Andrew Jackson, Randall K Ten Haken, Louis S Constine, Avraham Eisbruch, Søren M Bentzen, Jiho Nam, and Joseph O Deasy. Use of normal tissue complication probability models in the clinic. *International Journal of Radiation Oncology* Biology* Physics*, 76 (3):S10–S19, 2010.

Minesh Mehta, Rufus Scrimger, Rock Mackie, Bhudatt Paliwal, Rick Chappell, and Jack Fowler. A new approach to dose escalation in non–small-cell lung cancer. *International Journal of Radiation Oncology* Biology* Physics*, 49(1):23–33, 2001.

Arinbjörn Olafsson and Stephen J Wright. Efficient schemes for robust imrt treatment planning. *Physics in medicine and biology*, 51(21):5621, 2006.

Steven F Petit, Hugo JWL Aerts, Judith GM van Loon, Claudia Offermann, Ruud Houben, Bjorn Winkens, Michel C Öllers, Philippe Lambin, Dirk De Ruysscher, and André LAJ Dekker. Metabolic control probability in tumour subvolumes or how to guide tumour dose redistribution in non-small cell lung cancer (nsclc): an exploratory clinical study. *Radiotherapy and Oncology*, 91(3):393–398, 2009.

H Edwin Romeijn and James F Dempsey. Intensity modulated radiation therapy treatment plan optimization. *Top*, 16(2):215–243, 2008.

H Edwin Romeijn, Ravindra K Ahuja, James F Dempsey, Arvind Kumar, and Jonathan G Li. A novel linear programming approach to fluence map optimization for intensity modulated radiation therapy treatment planning. *Physics in Medicine and Biology*, 48(21):3521, 2003.

U.G.A. Sattler and W. Mueller-Klieser. The anti-oxidant capacity of tumour glycolysis. *International journal of radiation biology*, 85(11):963–971, 2009.

Lindsay I Smith. A tutorial on principal components analysis. *Cornell University, USA*, 51:52, 2002.

Marine Soret, Stephen L Bacharach, and Irene Buvat. Partial-volume effect in pet tumor imaging. *Journal of Nuclear Medicine*, 48(6):932–945, 2007.

CP South, Mike Partridge, and PM Evans. A theoretical framework for prescribing radiotherapy dose distributions using patient-specific biological information. *Medical physics*, 35:4599, 2008.

CP South, PM Evans, and M Partridge. Dose prescription complexity versus tumor control probability in biologically conformal radiotherapy. *Medical physics*, 36:4379, 2009.

Åste Søvik, Eirik Malinen, Øyvind S Bruland, Søren M Bentzen, and Dag Rune Olsen. Optimization of tumour control probability in hypoxic tumours by radiation dose redistribution: a modelling study. *Physics in medicine and biology*, 52(2):499, 2006.

Kenji Suzuki, Pingkun Yan, Fei Wang, and Dinggang Shen. Editorial: machine learning in medical imaging. *Journal of Biomedical Imaging*, 2012:2, 2012.

Florent Tixier, Catherine Cheze Le Rest, Mathieu Hatt, Nidal Albarghach, Olivier Pradier, Jean-Philippe Metges, Laurent Corcos, and Dimitris Visvikis. Intratumor heterogeneity characterized by textural features on baseline 18f-fdg pet images predicts response to concomitant radiochemotherapy in esophageal cancer. *Journal of Nuclear Medicine*, 52(3):369–378, 2011.

Iuliana Toma-Dasu, Johan Uhrdin, Alexandru Daşu, and Anders Brahme. Therapy optimization based on non-linear uptake of pet tracers versus linear dose painting. In *World Congress on Medical Physics and Biomedical Engineering, September 7-12, 2009, Munich, Germany*, pages 221–224. Springer, 2009.

Jan Unkelbach, Timothy CY Chan, and Thomas Bortfeld. Accounting for range uncertainties in the optimization of intensity modulated proton therapy. *Physics in medicine and biology*, 52(10):2755, 2007.

Manushka Vaidya, Kimberly M Creach, Jennifer Frye, Farrokh Dehdashti, Jeffrey D Bradley, and Issam El Naqa. Combined pet/ct image characteristics for radiotherapy tumor response in lung cancer. *Radiotherapy and Oncology*, 2011.

A. van Baardwijk, C. Dooms, R.J. van Suylen, E. Verbeken, M. Hochstenbag, C. Dehing-Oberije, D. Rupa, S. Pastorekova, S. Stroobants, U. Buell, et al. The maximum uptake of 18 f-deoxyglucose on positron emission tomography scan correlates with survival, hypoxia inducible factor-1α and glut-1 in non-small cell lung cancer. *European Journal of cancer*, 43(9):1392–1398, 2007.

Jason Van Hulse, Taghi M Khoshgoftaar, and Amri Napolitano. Experimental perspectives on learning from imbalanced data. In *Proceedings of the 24th international conference on Machine learning*, pages 935–942. ACM, 2007.

Ruud GPM van Stiphout, Guido Lammering, Jeroen Buijsen, Marco HM Janssen, Maria Antonietta Gambacorta, Pieter Slagmolen, Maarten Lambrecht, Domenico Rubello, Marcello Gava, Alessandro Giordano, et al. Development and external validation of a predictive model for pathological complete response of rectal cancer patients including sequential pet-ct imaging. *Radiotherapy and oncology*, 98 (1):126–133, 2011.

Barbara Vanderstraeten, Werner De Gersem, Wim Duthoy, Wilfried De Neve, and Hubert Thierens. Implementation of biologically conformal radiation therapy (bcrt) in an algorithmic segmentation-based inverse planning approach. *Physics in medicine and biology*, 51(16):N277, 2006a.

Barbara Vanderstraeten, Wim Duthoy, Werner De Gersem, Wilfried De Neve, and Hubert Thierens. [¡ sup¿ 18¡/sup¿ f] fluoro-deoxy-glucose positron emission tomography ([¡ sup¿ 18¡/sup¿ f] fdg-pet) voxel intensity-based intensity-modulated radiation therapy (imrt) for head and neck cancer. *Radiotherapy and oncology*, 79(3):249–258, 2006b.

Johan F Vansteenkiste, Sigrid G Stroobants, Patrick J Dupont, Paul R De Leyn, Erik K Verbeken, Georges J Deneffe, Luc A Mortelmans, Maurits G Demedts, et al. Prognostic importance of the standardized uptake value on 18f-fluoro-2-deoxy-glucose–positron emission tomography scan in non–small-cell lung cancer: an analysis of 125 cases. *Journal of Clinical Oncology*, 17(10):3201–3206, 1999.

Hubert Vesselle, Rodney A Schmidt, Jeffrey M Pugsley, Melissa Li, Steve G Kohlmyer, Eric Vallières, and Douglas E Wood. Lung cancer proliferation correlates with [f-18] fluorodeoxyglucose uptake by positron emission tomography. *Clinical Cancer Research*, 6(10):3837–3844, 2000.

S Webb and AE Nahum. A model for calculating tumour control probability in radiotherapy including the effects of inhomogeneous distributions of dose and clonogenic cell density. *Physics in medicine and biology*, 38(6):653, 1999.

Terry Windeatt. Ensemble mlp classifier design. In *Computational Intelligence Paradigms*, pages 133–147. Springer, 2008.

Marnix Witte, Georgy Shakirin, Antonetta Houweling, Heike Peulen, and Marcel van Herk. Dealing with geometric uncertainties in dose painting by numbers: Introducing the δvh. *Radiotherapy and Oncology*, 100(3):402–406, 2011.

Bradly G Wouters and J Martin Brown. Cells at intermediate oxygen levels can be more important than the" hypoxic fraction" in determining tumor response to fractionated radiotherapy. *Radiation research*, 147(5):541–550, 1997.

Lei Xing, Cristian Cotrutz, Sandeep Hunjan, Arthur L Boyer, Elfar Adalsteinsson, and Daniel Spielman. Inverse planning for functional image-guided intensity-modulated radiation therapy. *Physics in medicine and biology*, 47(20):3567, 2002.

Yong Yang and Lei Xing. Towards biologically conformal radiation therapy (bcrt): selective imrt dose escalation under the guidance of spatial biology distribution. *Medical physics*, 32:1473, 2005.

H Young, R Baum, U Cremerius, K Herhloz, O Hoekstra, AA Lammertsma, et al. Position paper. measurement of clinical and subclinical tumour response using [18 f]-fluorodeoxyglucose and positron emission tomography: review and 1999 eortc recommendations. *Eur J Cancer*, 35:1773–1782, 1999.