Combined Radiomics-Dosiomics for Radiation Pneuomonitis Prediction

by

Nitish Chopra

B.S., University of Wisconsin - Madison (2012)

August 2020

Submitted to the Department of Physics & Applied Physics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Physics - Medical Physics Option

at the

UNIVERSITY OF MASSACHUSETTS - LOWELL

Combined Radiomics-Dosiomics for Radiation Pneuomonitis Prediction

A Dissertation Presented
by

Nitish Chopra

B.S., University of Wisconsin - Madison (2012)

August 2020

Author .....
Department of Physics & Applied Physics
PhD Candidate (UML)

Certified by .....
Erno Sajo, PhD
Professor (UML)
Committe Co-chair

Certified by .....
Raymond Mak, MD
Brigham and Women's Hospital
Committee Co-chair

Certified by .....
Gregory Sharp, PhD
Associate Professor (MGH, HMS)
Committee Member

Certified by .....
Joyita Dutta, PhD
Associate Professor (UML)
Committee Member

Certified by .....
Wilfred Ngwa, PhD
Adjunct Professor (UML)
Committee Member

Combined Radiomics-Dosiomics for Radiation Pneuomonitis Prediction

by

Nitish Chopra

Abstract of a dissertation submitted to the faculty of the Department of Physics & Applied Physics in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Physics - Medical Physics Option

at the

UNIVERSITY OF MASSACHUSETTS - LOWELL

August 2020

Dissertation Supervisors:

Dr. Erno Sajo, Professor
Department of Physics & Applied Physics
University of Massachusetts-Lowell

Dr. Raymond Mak, Associate Professor
Harvard Medical School
Brigham & Women's Hospital, Boston

Abstract

Radiation Pneumonitis (RP) is a commonly diagnosed morbidity and major dose-limiting factor for patients undergoing radiation therapy treatment for advanced stage lung cancer. Improving the local-control necessitates safely escalating the radiation dose to the tumor while minimizing the normal tissue toxicity. We present a novel radiodosiomics machine-learning framework to evaluate the combined prognostic power of pre-treatment radiomics (CT) and dosiomics (TPS Dose Image) features to predict radiation pneumonitis (RP) in a large cohort of 701 locally advanced non-small cell lung cancer (NSCLC) patients. Dose-based ROIs (ROI-V5 & ROI-V20) registered to the CT simulation were employed for feature extraction. Radiomics and Dosiomics features were computed for the bilateral and ipsilateral dose-based pulmonary ROIs for patients that completed curative treatment (median dose=64 Gy). Features were evaluated for robustness before selection was performed. A clinical model was also curated for comparison that included conventional metrics such as patient age, gender, smoking status, pack years, concurrent chemotherapy status, treatment modality, and performance status as well as DVH-based metrics namely V20, MLD, and V5. Univariate and multivariate analysis was stratified by radiation modality. A robustness study was conducted to ensure feature stability before using them for model-building. Our results demonstrate that the *Radiomics+Dosiomics* random forest classifier provides statistically significant ($p<0.05$) augmentation to the *Clinical* model for RP prediction. The ROI-V5 bilateral *Radiomics+Dosiomics+Clinical* (AUC=0.81) model was the highest performing model and statistical better than the *Clinical* (AUC=0.73) model. Moreover, this statistically significant finding was observed across both ROIs and lung-sides (bilateral vs ipsilateral). While the combined radiomics and dosiomics approach performance was stronger for IMRT dataset, the *Radiomics+Dosiomics* model was extremely effective at predicting RP for 3DCRT and significantly ($p\leq0.05$) better than the *Clinical* model. The RP risk assessment framework presented here, if clinically implemented, would allow a radiation oncologist to up or down-regulate dose prescriptions based on individual risk that is based on anatomical baseline differences, the dose distributions and the interactions thereof.

Acknowledgments

I would like to acknowledge significant contributions made by some wonderful people who I leaned on during my PhD. First and foremost, I would like to thank my advisor and medical physics graduate program director, Dr. Erno Sajo, at the University of Massachusetts-Lowell for trusting me with the graduate position. I will be forever grateful for the opportunity.

I want to express my deepest gratitude towards my committee co-chair and research advisor, Dr. Raymond Mak, for counseling me and ensuring that our research efforts were always grounded in clinical realities. I would also like to acknowledge my mentors at Massachusetts General Hospital (MGH), Greg Sharp and Bruce Crawford, for always taking time to train me during my tenure as a Medical Physics Assistant, a role that allowed me to build on the didactic knowledge in a highly clinical setting while simultaneously giving me the flexibility to undertake PhD research.

Additionally, I would like to recognize various members of the MGH radiation oncology family - Jennifer Pursley, Mandar Bhagwat, David Craft, Clemens Grassberger, Chris Cotter, Caitlin Finley and Daniel Shinnick for cheerleading my successes and challenging my research ideas when necessary.

Lastly and most importantly, I would like to thank my parents, my partner, and my labradoodle for their unending support, patience and love.

Contents

1 Introduction **1**

1.1 Non Small Cell Lung Cancer and Radiotherapy 2

1.2 Radiation Pneumonitis 3

1.3 Radiomics 7

1.4 Dosiomics 9

2 Methods **11**

2.1 Dataset 12

2.2 Image Processing and Pulmonary Volume Generation 12

2.3 Feature Extraction 13

2.4 Feature Robustness 14

2.5 Feature Selection 15

2.6 Clinical Predictors 16

2.7 Statistical Analysis 17

2.8 Class Imbalance 18

3 Univariate Results **20**

3.1 Clinical Characterstics 21

3.2 Clinical Predictors of Radiation Pneumonitis 21

3.3 Bilateral Features 22

3.4 Ipsilateral Features 23

4 Logistic Classification Models **25**

4.1 Bilateral Models 26

4.2 Ipsilateral Model 27

5 Random Forest Classification Models **29**

5.1 Bilateral Models 30

5.2 Ipsilateral Model 31

6 Discussion **33**

7 Conclusion **37**

A Figures **39**

B Tables **57**

List of Figures

A-1 Radiation Modality as a function treatment start date. Introduction of IMRT at BWH. High risk patients elevated to IMRT instead of 3DCRT initially . . 40
A-2 Machine Learning Architecture for RP Prediction 41
A-3 Radiation Modality 42
A-4 Treatment Modality 42
A-5 Univariate predictive perform of clinical predictors for the pooled dataset and stratified by radiation modality 43
A-6 Univariate predictive performance of the the top 5 mRMR selected bilateral ROI-V20 Radiomics (left) and Dosiomics (right) features 44
A-7 Univariate predictive performance of the the top 5 mRMR selected bilateral ROI-V5 Radiomics (left) and Dosiomics (right) features 45
A-8 Univariate predictive performance of the the top 5 mRMR selected ipsilateral ROI-V20 Radiomics (left) and Dosiomics (right) features 46
A-9 Univariate predictive performance of the the top 5 mRMR selected ipsilateral ROI-V5 Radiomics (left) and Dosiomics (right) features 47
A-10 Random Forest performance of bilateral ROI-V20 features compared against the clinical model using permutation test (** - $p \leq 0.01$, * - $p \leq 0.05$). Radiodosiomics model augments the clinical model but the gains are more noticeable in the IMRT subgroup 48
A-11 Random Forest performance of bilateral ROI-V5 features compared against the clinical model using permutation test (** - $p \leq 0.01$, * - $p \leq 0.05$). Radiodosiomics (Radiomics+Dosiomics) model augments the clinical model. Absolute performance is especially stronger for the IMRT subsample 49

A-12 Random Forest performance of ipsilateral ROI-V20 features compared against the clinical model using permutation test (** - p≤0.01, * - p≤0.05). Radiodosiomics model augments the clinical model but the gains are more noticeable in the IMRT subsample 50
A-13 Random Forest performance of ipsilateral ROI-V5 features compared against the clinical model using permutation test (** - p≤0.01, * - p≤0.05). Radiodosiomics (Radiomics+Dosiomics) model augments the clinical model. Absolute performance is especially stronger for the IMRT subsample 51
A-14 Logistic classifier performance of the bilateral ROI-V20 radiomics, dosiomics and the combined models plotted along with the clinical model 52
A-15 Logistic classifier performance of the bilateral ROI-V5 radiomics, dosiomics and the combined models plotted along with the clinical model 53
A-16 Logistic classifier performance of the ipsilateral ROI-V20 radiomics, dosiomics and the combined models plotted along with the clinical model 54
A-17 Logistic classifier performance of the ipsilateral ROI-V5 radiomics, dosiomics and the combined models plotted along with the clinical model 55
A-18 Locoregional failure rates. A significant number of patients who were treated with 3DCRT and IMRT reccured within 1 year 56
A-19 The process of generating ROI-V5 by using the anatomical (CT) and dose distribution (TPS). The tumor subtracted pulmonary structures were overlayed with the dose distribution to retain a binary pulmnonary volume that received ≥ 5 Gy 56

List of Tables

B.1 Breakdown of 1967 Radiomics & Dosiomics features listing the enabled feature classes and corresponding number of features extracted from the different image types i.e Original (unfiltered) image, Laplacian of Gaussian (LoG) filtered image, and Wavelet filtered image 58

B.2 Number of robust radiomics and dosiomics features (ICC≥0.80). Radiomics robustness was evaluated on the RIDER Test-Retest dataset (n=22) and the dosiomics robustness was evaluated on the dose images of the excluded cases (n=102). 58

B.3 Feature Datasets 58

B.4 Patient Characterstics, Treatment Information, and Outcomes 59

B.5 Dosimetric Characterstics 60

B.6 Encoding type for all the clinical variables considered in the *Clinical* model . 60

B.7 Test AUC for bilateral ROI-V20 random forest models iterated 100 times into 70% train, 30% test with 10-fold cross validation and SMOTE resampling to address class imbalance. 61

B.8 Test AUC for bilateral ROI-V5 random forest models iterated 100 times into 70% train, 30% test with 10-fold cross validation and SMOTE resampling to address class imbalance. 61

B.9 Test AUC for ipsilateral ROI-V20 random forest models iterated 100 times into 70% train, 30% test with 10-fold cross validation and SMOTE resampling to address class imbalance. 62

B.10 Test AUC for ipsilateral ROI-V5 random forest models iterated 100 times into 70% train, 30% test with 10-fold cross validation and SMOTE resampling to address class imbalance. 62

B.11 Test AUC for bilateral ROI-V20 logistic classification models iterated 100 times into 70% train, 30% test with 10-fold cross validation and SMOTE resampling to address class imbalance. 63
B.12 Test AUC for bilateral ROI-V5 logistic classification models iterated 100 times into 70% train, 30% test with 10-fold cross validation and SMOTE resampling to address class imbalance. 63
B.13 Test AUC for ipsilateral ROI-V20 logistic classification models iterated 100 times into 70% train, 30% test with 10-fold cross validation and SMOTE resampling to address class imbalance. 64
B.14 Test AUC for ipsilateral ROI-V5 logistic classification models iterated 100 times into 70% train, 30% test with 10-fold cross validation and SMOTE resampling to address class imbalance. 64

Chapter 1

Introduction

1.1 Non Small Cell Lung Cancer and Radiotherapy

Global statistics show that lung cancer is the most commonly diagnosed cancer and is the leading cause of cancer deaths [65]. According to a 2018 study, approximately 18.4 % of all cancer deaths were attributed to lung cancer [6]. Mortality due to lung cancer is expected to climb up 10 million in the next 15 years [66]. Non-Small cell lung cancer (NSCLC) is one of the primary type of malignancies in patients presenting with lung cancer. Nearly 80% of all patients with pulmonary malignancies present with NSCLC and a high majority of them are locally advanced at the time of diagnosis. It is estimated that NSCLC accounts for nearly 23% and 15% of all cancer deaths in men and women respectively [44]. Even for early stage resectable NSCLC, around 30-55% of patients will develop disease recurrence within the first 5 years of surgery [57]. Advanced stage NSCLC have a markedly higher risk of recurrence and a higher propensity for distant recurrence [41].

While the outcomes in terms of overall survival and loco-regional control have been trending in the positive direction with the advent of modern and more safer radiation modalities, the prognosis is far from satisfactory given the high incidence of NSCLC. As such, there is still a large room for improvement. In a meta-analysis of published studies, median overall survival (OS) was found to be 28.7 months and locoregional control was 65% in the conventional treatment arm [67].

Active research is ongoing into newer treatment modalities so as to affect better outcomes. Work is being conducted on various fronts including search for check point inhibitors in immunotherapy as well alternative fractionation regimens. Both of these options have yet to overcome significant roadblocks. While an improvement towards overall survival was observed in some immunotherapy studies, generally the efficacy results are mixed [66].

Majority of the available options treating NSCLC include radiation therapy (RT). Historically, RT was employed as part of the sequential chemo-radiotherapy framework. Concommittant chemo-radiotherapy became the mainstay since it was shown to achieve better outcomes in prospective radiomized control trials compared to sequantial chemo-RT [67]. The current standard of care for locally advanced NSCLC is concomittant chemo-radiotherapy with hyperfractionated 60 Gy with two cycles of platinum-based chemotherapy. However, platinum-based chemotherapies come their own associated morbidities and risks. Moreover, the efficacy of platinum based therapies have been shown to correlate with gene mutation

status, specifically, single nucleotide poumorphisms (SNP) [42]. More recently, stereotactic body radiation therpay (SBRT) has seen successful clinical adoption as it aims to deliver much higher doses to the tumor over a much smaller number of fractions. But the higher dose also decreases the margin of error and normal tissue toxicity becomes an even larger consideration.

A meta-analysis by Mauguen et al. revealed an overall survival benefit for patients treated with alternative fractionation as opposed to conventional fractionation. Hyperfractionation and hypofractionation may contribute to an increase in the effective total dose. As such, adaptations to both fractination regimens are being explored to achieve better outcomes. However, trade-offs between high therapeutic dose vs normal tissue toxicity must be carefully balanced.

1.2 Radiation Pneumonitis

Despite the advances that have been made in terms of conformal radiation therapy, Radiation induced lung toxicity (RILT) is still a major consideration while treating patients with lung cancer [26]. The effectiveness of delivering radiation therapy has been largely limited by normal tissue injury [30]. The manner in which these sideffects manifest can vary greatly. They may be short-lived or long-lasting. Some of these conditions are associated with edema, epithelial degeneration, invasion of alveoli by bronchial epithelium, endothelial sloughing, disruption of microvasculature, and atelectasis [32]. Some of these reactions may be reversable but others fail to repair, leading the tissue into a dysregulated state that can manifest as fibrosis or radiation pneuomnitis [30].

Radiation Pneumonitis (RP) is an iatrogenic sideffect seen in patients treated for thoracic malignancies with radiation therapy (RT). The incidence of symptomatic RP ranges from around 10-30% [42]. Time to onset for RP tends to be anywhere from weeks to months following radiation therapy but most cases are reported within the first 8 months with patient presenting dyspnea on exertion, non-productive cough, and hypoxemia as the most common symptoms [30]. Physical exam findings may include pleural friction rub and moist rales [25].

The risk of developing RP can be attributed to either patient-related factors or treatment/clinical factors. The manifestations of RP can also be complicated by prior lung

disease. In a meta-analysis published by Vogelius et al., advanced age and pulmonary comorbidities were shown to be significantly associated with the risk of developing RP [60]. Surprisingly, in the same study, ongoing smoking was shown to be protective which may be explained by lower inflammatory reaction to radiation among smokers.

Pre-existing pulmonary issues are fairly common in patient population undergoing radiation treatment for NSCLC. Generally, these pulmonary conditions are classified into two broad categories - obstructive or restrictive. Conditions like Asthma and chronic obstructive pulmonary disease (COPD) fall in the obstructive category and emphysema and interstitial lung disease (ILD) fall in the restrictive category [30]. The effect of COPD on RP is mixed. On the other hand, ILD is strongly associated with increased risk of RP. In a large retrospective analysis of nearly 504 patients treated with sterotactic body radiation, grade ≥ 3 RP was higher in the ILD group (32 %) compared to the group without ILD (2 %, $p< 0.001$) [2]. Similar correlations with ILD for patients who developed RP were found in another study of 651 patients treated with RT; 78 % of patients that developed RP were found to have pre-existing ILD [46].

The COVID-19 crisis is exposing the population to a range of debilitating chronic conditions. Predicting a patient's risk of RP post-SARS-nCOVID-19 pandemic is especially relevant due to the nature of the disease. One of the major manifestations of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus is that it directly affects the repiratory system of the host [52]. While it is too soon to know exactly how the comorbidities brought on by COVID-19 will affect a patient's risk of developing RP, it is certain that a significant portion of our future patient populations will have previous COVID exposure. Therefore, it will be imperative to gather data and develop models that are able to easily incorporate COVID-19's effect on RP risk.

As stated in 1.1, while early stage lung cancer can be treated definitively with either surgery or radiation alone, management of advanced lung cancer involves some combination of chemotherapy and radiation. The two modalities are known to have snergistic effect but the sequencing of these two modalities has been a matter of debate with mixed findings [30]. Currently, concurrent treatment is the standard of care. Sequential treatments may be considered based on elevated complexicity of a particular case, owing to large tumor volumes and unmet dosimetric cutoffs or poor performance status. Some studies have suggested an increased risk of RP for sequential chemoradiation (CRT) [61], however, the data maybe

biased as the CRT patients tend to be higher risk with poorer performance status [30].

More recently, immunotherapy has shown great promise in the management of NSCLC. Results from various studies have proven immunotherapeutic agents to be superior to standard chemotherapy [50, 28, 5]. However, the toxicities associated with immunotherapy are not fully understood yet and the we are just uncovering the surface. Nevertheless, ILD and immunotherapy are already shown to be strongly correlated in large retrospective trial of over 1800 patients [18]. Another recent analysis of published studies found higher RP incidence in patients treated with immunotherapy when compared to other drugs used in treating NSCLC [13, 27, 53, 24, 33, 40]. Since RT has been shown to prime the immune system for a response from immunotherapy, the synergistic increase in RP risk may need to also be considered.

RT is an integral part of the standard-of-care management of LA-NSCLC. RP remains a significant barrier to radiation dose escalation in order to achieve adequate local control in the treatment of locally advanced lung cancer [42]. Radiation Induced Lung damage (RILD) may be attributed to radiation toxicity to healthy tissue as well inflammatory changes that follow. In an effort to minimze the risk of radiation toxicity to normal pulmonary tissue, dosimetric cutoffs were proposed in late 1990s and have been studied extensively over the last two decades. Radiation dose to lung volumes has been correlated with risk of pneuomonitis. DVH-based metrics such as volume of lungs receiving > 20 Gray (Gy; V20), > 5 Gy (V5) and mean lung dose (MLD) have been shown to associate positively the risk of developing correlated with the risk of developing RP [23, 37, 62]. Corresponding dosimetric cutoffs, specifically lowering V20 and V5, to lower the risk of RT-induced RP, were proposed and are currently used to evaluate treatment plan quality in the clinic. Most studies have validated V20 and MLD as the primary parameters to be controlled to lower RP risk, in addition to various other DVH derived volumetric parameters. However, these additional metrics are often collinear i.e if one increase, the other one increases too, and too often the difference in there predictive power is minimal. Additional attempts to use conditional metrics that control for the dose to the heart and lungs have been put forward [59, 22] but follow-up studies have failed prove its usefulness [63]. As a result, the quest for more robust RP predictors has continued. These dosimetric cutoffs while useful in clinic, do not take into account a continuum of risk of normal tissue toxicity which may be determined by the interaction between dose distribution, patient based baseline anatomical differences seen on

the CT, as well as other underlying biological factors in the patient [42]

Radiation treatment technique also has significant impact on the RP risk. 3D-Conformal Radiation Therapy (3DCRT) use to be the mainstay for the management for LA-NSCLC. Compared to 3DCRT, newer image modulated radiation therapy (IMRT) provides better conformality, better target coverage and is highly effective at avoiding adjacent organs at risk (OARs) thereby enhancing the therapeutic ratio. In a recent prospective Phase 1 study of patients with LA-NSCLC undergoing definitive IMRT with corresponding 3DCRT plans, a decrease in normal tissue exposure for the IMRT cohort was observed [4]. The strongest evidence that propelled IMRT as the standard of care treatment for LA-NSCLC was provided form a secondary analysis by Chun et al. of RTOG 0617, a trial in which patients were treated with concurrent CRT with or without cetuximab to a dose of 60 vs 74 Gy [10]. Since the choice of RT technique was not randomized, the IMRT group were patient with larger tumors and higher complexicity. Even so, univariate and multivariate analysis showed that the patients in the IMRT group had significantly lower occurrence of Grade 3 or higher RP. Despite the improvements provided by IMRT, the dosimetrics variables used in clinic for treatment plan evaluation condense a three dimensional distribution to a single metric and may not capture all the potentially useful spatial information. Therefore the quest for more robust RP biomarkers has proceeded on.

Another major factor in studying pneumonitis is the variation in grading across physicians and clinics. Some of the commonly used standard for RP grading are: *a) Common Terminology Criteria for Adverse Events (CTCAE) b)* Radiation Therapy Oncology Group (RTOG) *c)* European Organization for Research and Treatment of Cancer (EORTC) *d)* Southwest Oncology Group (SWOG) *e)* Eastern Cooperative Oncology Group *f)* World Health Organization. Our work utilized the the CTCAEv.4 criteria for RP grading. Finding consistency in the grading of RP can be a challenge in itself owing to various evolving standards. These standards attempt to encompass the various clinical, anatomical and radiological factors present in a subject following radiotherapy. Newer deep learning based models could help consolidate these available grading schemes so that there is little intra-intstituition and inter-instituition variability in the diagnosing the presence and extent of RP.

1.3 Radiomics

Radiomics is a recently developed technique that aims to address the drawbacks of conventional outcome models by using high dimensional mineable features extracted from large number of data-characterization algorithms [38, 36]. Historically, tumor characterization is based on invasive biopsies or visual assessments of medical images. Biopsies may give an incomplete understanding of the tumor characteristics as it is derived via random spatial sampling of the tumor. Moreover, malignant tumors are known to be spatially and temporally heterogeneous as result of regional variations in metabolism, gene expression, vasculature and oxygenation [43, 45, 9, 19]. In addition, as state earlier, medical image evaluation can itself be subjective and physician dependent. Visual assessments of medical images can suffer from large intra-observer and inter-observer variability [17]. The radiomics framework aims to eliminate these weaknesses by standardizing the evaluation process rooting it in quantitative, computer-aided decision making.

The advances in medical imaging have made possible non-invasive sampling of the tumor and its microenvironment. The phenotypic patterns of anatomy, i.e tumor, as seen across imaging modalities has been associated with underlying cellular and genetic factors [16]. CT imaging features have been found to be associated with anaplastic lymphoma kinase (ALK) mutations in lung cancer patients [64]. Another study found significant correlation of CT images with "gross appearance of intra-tumoral vascularity" and "well-defined tumor boundaries". Similar associations with underlying physiological factors have been presented in magnetic resonance imaging (MR) and in PET imaging data. Aerts et al. showed using multiple external validation data, that radiomics signatures could be used for outcome prediction and clinical decision making [1].

CT-based features have been particularly successfully implemented to predict various endpoints for NSCLC datasets. In a study of adenocarcinoma patients, Coroller et al. were able to use pre-treatment CT-based signature as a prognostic biomarker for distant metastatsis [12]. Additional outcomes including disease-free survival, overall survival, pathological complete response, and local recurrence have been successfully studied using CT-based radiomics [11, 29].

Delta-radiomics is another application which has been used for outcome modeling. As the name suggests, delta-radiomics uses features that describe the change in the feature

value over the course of a given treatment (RT or CRT). Studies have suggested that such an approach can improve prognostic value of outcome models. This approach has primarily been applied on CT and PET imaging data. Fave et al. employed mixed-effects models to identify markers that decrease significantly over the course of treatment thereby showcasing the value-added nature of delta-feature analysis to model overall survival, distant metastasis and local recurrence. Carvalho et al hypothesized that the FDG-PET metabolic uptake patterns may allow for early response assessment over the course of treatment. In a prospective dataset of 54 local advanced NSCLC, their approach provided benefit of timely assessment of response to radiotherapy when compared to human-based inference [7].

While the initial impetus of radiomics was based around prognostication and outcome modeling, CT as well PET-based radiomics have shown promise in modeling radiation-based toxicities especially RP. Traditional models for predicting RP models lack the the ability to adequately consider the baseline differences in patient lung function and radiosensitivity [34]. Historically, RP studies have relied on radiologist-examined semiquantitaive scoring of interstitial lung disease which does not capture hetrogeneity in lung function across the lung population [34]. Recent advances in image analysis have been focused around shifting from a qualitative to a quantitative analysis of CT data [38]. In addition to conventional clinical characterstics such as tumor size and shape, CT features are able to capture the remodeling of normal pulmonary tissue as a result of radiation therapy. In a study by Cunliff et al, curated "delta-features" from pre and/or post CT were shown to be moderately associated with RP (grade >2) in a small cohort of patients that were treated for esophageal cancer using "serial delta-radiomics" approach [15]. In a follow-up prospective study, Cunliff et al ([14] correlated the change in CT feature values with radiation dose, thereby showcasing the ability of radiomics to measure patient lung tissue reaction to RT and assess RP development. The primary issue with a delta-feature based model is that it does not allow to any plan adaptation to minimize the RP risk. In an ideal scenario, normal tissue toxicity model should be able to aid physicians pre-treatment in regulate radiotherapy dose and/or other treatment prescriptions. As such, the effort must be focussed on an approach that enables a robust pre-treatment RP risk evaluation.

1.4 Dosiomics

The success of radiomics and machine-learning based quantitative analysis of medical imaging has precipitated new ways in which other data streams could be utilized for outcome modeling. Specifically in regarding to modeling normal tissue toxicity modeling, researchers have adapted radiomics-type analysis to features extracted using 3D dose distributions (Dosiomics) to supplement or outperform DVH-based models. As mentioned in SECTION, the idea of using DVH metrics assumes that these metrics describe most, if not all, of the useful information that a physicians needs to minimize the dose to OARs and normal tissue toxicity. DVH metrics are certainly more intuitive to understand and regulate. However, as the treatment planning transitions to into semi and fully automated algorithms, effort are also undergoing to uncover predictors to model toxicity through a more multivariate lens. Computerized engines could easily be trained to optimize such metrics that are not limited by easy human interpretation.

One of the most promising implementation of Dosiomics was by Gabrys et al, for xerostomia risk assessment in comparison with normal tissue complication probability (NTCP) models based on mean radiation dose to the parotid glands. Their analysis, they used dosiomics features in addition to radiomics and NTCP predictors to build logistic, random forest (RF) and support vector machine (SVM) classifiers and were able to infer that dose-based shape descriptors were beneficial for xerostomia prediciton in highly conformal radiotherapy patients [20]. In another recent work by Rossi et al, inclusion of dosiomics features to DVH based metrics was shown to provide statistically significant improvement in prediction of gastrointestinal and genitourinal toxicity rates in prostate cancer patients [51].

Moderately successful attempts to apply dosiomics for RP prediction have been made. In a study of 70 NSCLC patients treated with volumetric arc therapy (VMAT) radiation, Liang et al. demonstrated that the dosiomics features have approximately equivalent predictive power as NTCP predictors and that the spatial dosiomics features are able to able to predict RP. In a follow-up study, the same group implemented convolutional neural networks on the 3D dose ditributions with modest performance to predict RP. Generalization of such an approach on a larger, more representative dataset is unknown and needs to be evaluated [39].

However, such an approach of using dosiomics features alone or even in combination with NTCP predictors does not account for patient-specific anatomical data that is readily available in the clinic. In modern day radiation oncology clinics, every patient undergoes a pre-treatment CT simulation for treatment planning purpose. Devising an approach that combines anatomical and dose-based predictors could allow for a more comprehensive RP risk evaluation. In addition, as more studies understand the detrimental effects of COVID-19 on the pulmonary tissue and immune system, it is all the more important to invent methods that leverage CT data as it may inform the model. The motive behind the work presented hereforth was to implement a new methodology that combines radiomics and dosiomics predictors extracted from the pulmonary volumes so as to robustly predict RT in a large dataset of 701 NSCLC patients.

Chapter 2

Methods

2.1 Dataset

For our Institutional Review Board (IRB) approved prospective study, we considered a large cohort of 701 NSLSC patients with locally advanced stage 3 lung cancer that were treated with radiation therapy at Brigham and Women's Hospital, Boston, MA between 2001 and 2014. Patients who did not complete their courses (n=63) or had severe CT artifacts (n=42) were excluded from the study. Table B.4 and Table B.5 list the clinical and dosimetric characterstics of the remaining 596 patients. Nearly three-fourth (Figure A-3) of the patients were treated with 3DCRT and last fourth were treated with conformal IMRT/VMAT doses. Nearly 59% of the patients (Figure A-4 were treated with concurrent chemotherapy, around 24% were treated with neoadjuvant therapy (chemo + radiation + surgery), around 10% of the patients were post-surgery cases that were also included in our dataset. The remaining 6% were treated with definitive radiation therapy. Based on histological data, most of the patients were treated for adenocarinoma, followed by squamous cell carcinoma (Figure B.4).

The cohort analyzed in this work spans more than a decade ranging from year 2003 - 2014. 3DCRT was the standard-of-care way of delivering curative treatment for patient with thoracic tumors until the advent of image modulated radiation therapy (IMRT). These clinical advancements are reflected in our cohort which includes patients that were treated with 3DCRT (75%) as well as IMRT (25%). Figure A-1 depicts the radiation start date for each of the patients. IMRT was commissioned at BWH in 2009 and the patients that received the newer modality were chosen by the physicians and not randomized. The RP grading for this study was based on the independent ratings of two physicians that utilized the CTCAEv.4 criteria.

2.2 Image Processing and Pulmonary Volume Generation

The treatment planning CT, planned dose as well as the tumor and organ volumes as delineated by the physicians were extracted from the treatment planning software (TPS). All the Digital Imaging and Communications in Medicine (DICOM) files were converted to a more portable Nearly Raw Raster Data (NRRD) format using plastimatch [56]. We used two seperate RT plan dose-based pulmonary regions of interests (ROIs) for each patient (ROI-V5 and ROI-V20). In other words, for each patient the pulmonary ROIs used for

feature extracted are tied to the dose-distribution that they eventually received during the treatment. The general workflow for the pulmonary volume generation is shown in Figure A-19. Irradiated volume $V_{\text{dose (Gy)}}$ is a commonly used metric to gauge the toxicity risk of a thoracic radiation therapy treatment plan. Both $V_{20\text{ Gy}}$ and $V_{5\text{ Gy}}$ are associated with radiation based toxicity in lung cancer. ROI-V5 corresponds to the region around the tumor that received ≥ 5 Gy and ROI-V20 corresponds to the region around the tumor that received ≥ 20 Gy.

For each patient, CT registered TPS dose distribution was overlayed to extract out the two ROIs by first subtracting the internal tumor volume (ITV) from the lung. For cases that did not have an ITV, gross tumor volume (GTV) was used. Clinical tumor volume (CTV) was used for a subset of surgical cases that are also included in the dataset. The intersection of respective V5 and V20 dose and pulmonary contours was used to extract the ROIs. The ROIs were also labelled according to the lung sidedness so as to eventually be able to develop bilateral and ipsilateral RP models.

2.3 Feature Extraction

Various open-sourced toolkits are available to extract radiomics features. We adopted the PyRadiomics open-sourced package for feature extraction. The PyRadiomics platform can extract features from various medical images (such as CT, PET, MRI) [58]. The images are first pre-processed which involves basic imaging processing such as isotropic voxel resampling. Various imaging filter can be enabled and applied on the original (unfiltered) image via a customization parameter file. Enabling imaging filters allows the extracted features to capture different textures. Table B.1 shows the enabled feature classes and the corresponding number of features in each class. In addition to the unfiltered image, the same set of features were extracted for filtered images. Three image types were enabled *a) Original (Unfiltered) b) Laplacian of Gaussian (LoG) Filtered Image c) Wavelet Filtered Image*

Laplacian of Gaussian filter is an edge enhancement filter which emphasizes the areas of grey-level change. Parameter σ is used to define how coarse the emphasized texture should be with a low sigma value emphasizing finer textures, where as a high sigma value emphasizes coarser textures. Features were computed for five sigma values $\sigma \in$ [1-5]. Two-level wavelet filter was enabled with each level yielding 8 different decompositions per level.

Barring the original image itself, 15 unique wavelet decompositions were used for wavelet feature extraction.

The ROI-V5 and ROI-V20 were used to extract 1967 radiomics features from the CT scan and 1967 dosiomics features from the TPS dose distribution. Voxels were resampled to 3 mm in each dimension prior to feature extraction to standardize the voxel spacing for all the cases in the dataset. For radiomics features a voxel intensity binwidth of 25 hounsfield units (HU) was used for textural features. For dosiomics features, a dose intensity of 1 Gy was used for textural features. Features were computed for the bilateral and ipsilateral pulmonary ROIs. Once the features were calculated the results are written to the comma-separated file and ready for further analysis. In summary, eight seperate dataset feature files (Table B.3) were analyzed using the workflow depicted in Figure A.

2.4 Feature Robustness

The quantitative analysis of CT and Dose Images involves converting the imaging data into mineable data by extracting a large number of statistical features. Before using these features as potential biomarkers in the model, their robustness must be evaluated. Many of these features are often found to be unstable between imaging scans acquired within weeks-even minutes of each other (Tixier 2012).

In a study by Balagurunathan et al. in which repeated scans (Test and Re-Test scans) were taken 15 minutes apart, only 30% of the considered features were found to be robust (ICC $>$ 0.90) The robustness was defined by rater variability statistic- Intra-Class-Correlation coefficient (ICC). This suggests that a large number of the features may be unreliable and the robustness must be evaluated before moving forward with the step of model building. To our knowledge, the repeatability of dosiomics features has not been investigated. Understanding the stability of RT Dose Image features will be important for future dosiomics applications.

Both Radiomics and Dosiomics features were individually evaluated for robustness using the intra-class correlation coefficient (ICC). The analysis was performed using the R package irr [21]. In order to evaluate the robustness of radiomics features, we considered a Test-Retest dataset [68]. 22 scans were taken 15 minute apart using the same CT scanner and imaging protocol. The tumors were delineated by the same physicians. ROI-V5 and ROI-

V20 pulmonary volumes were extracted for each of these cases. A 25 mm expansion around the tumor was used for ROI-V20. The DVH metrics showed that the mean V5 for the dataset was nearly 73% larger than V20 for our cohort. Therefore, a 44mm expansion was used to create the ROI-V5 for the Test-Retest. Radiomics feature robustness was assessed on a Test-Retest Rider dataset using the methodology described by Chintan et al [47]. Feature with ICC ≥ 0.80 were retained for further analysis.

A dosiomics robustness study was devised using the planned dose distributions for patients that were excluded from the analysis due to incomplete prescription or CT artifacts. ROI-V20 and ROI-V5 were created for these patients. Features were computed for ROIs by shifting them by $\pm$3mm in x,y,z directions to replicated patient setup uncertainity. Only features with ICC ≥ 0.80 in all six directions were retained for further analysis. Dosiomics robustness was stratified by radiation modality (3DCRT & IMRT). Table B.2 lists the number of radiomics and dosiomics features that were found to be robust according the methodology described above. Out of the 1967 radiomics features that were extracted, 1483 ($\sim$75%) ROI-V5 features and 1336 ($\sim$68%) ROI-V20 features were found to be stable. For the dosiomics study, the stability analysis was stratified by radiation modality. For the pooled dataset, 1222 ($\sim$62%) ROI-V5 features and 1185 ($\sim$60%) ROI-V20 dosiomics features were found to be stable. As expected, these statistics were largely maintained for the 3DCRT dataset owing to the large number of cases. Interestingly, a slightly lower number of IMRT features were found to be robust for the ROI-V5 and ROI-V20 features (Table B.2.

2.5 Feature Selection

The PyRadiomics feature extraction suite generates a large number of features from the input images. Not all features, however, provide additional useful information for model building. A high majority of these features tend to be highly correlated.

Traditional machine-learning approaches are different from deep learning methods in that the features can be hand engineered by using sophisticated algorithms. Deep learning methods, such convolution neural nets are able to peform this step implicitly while focussing on a single parameter called the loss function. Such an approach is suitable when a well-performing model is the end goal. In many applications, it may be important to explicitly identify the variables that contribute to a good performing machine learning model. Our

approach assumes that predicting RP is as important as learning about the building blocks of the model.

Feature selection or variable selection is the process used in machine-learning to select a subset of input variables based on certain metrics that maximize the predictive performance. In addition to getting insight into the input variables, feature selection methods also help avoid or minimize overfitting and provide fast and more cost-effective models [54]. Feature selection methods may fall under two broad categories dictated by the kind of dataset available - *a) Supervised Feature Selection* or *b) Unsupervised Feature Selection*

Supervised learning is when the classifier are trained on ground truth labels. Our work is operating under this framework and using the binary RP labels that based on physicians grading of RP. Unsupervised learning involves training a model on variables that are extracted via a clustering analysis of a multi-dimensional variable space when the ground truth labels are not available or possible.

Feature selection methods can also be categorized under three broad categories based on how they are implemented in the machine learning workflow.*a) Filter methods*, *b) Wrapper methods*, and *c) Embedded methods*. For our study, we used Minimum Redundancy and Maximum Relevance (mRMR) feature selection method which is filter-based supervised feature selection framework. Given a set of features and target labels, mRMR selects the relevant features based on F-statistic while minimizing correlated features using Pearson correlation coefficient [49]. The feature selection was carried out using the survcomp package in R statistical language [55]. The mrmr.cindex utility generates a ranking based on the concordance index of the variable given a binary outcome vector. For each of the eight datasets mentioned in B.3, the top five features were retained for the model building.

2.6 Clinical Predictors

In order to compare the performance of our approach with published clinical predictors of RP, several patient and treatment related variables were incorporated into a clinical model. The patient and treatment related variables included age, gender, smoking, pack years, concurrent chemotherapy status, treatment modality, and performance status were considered for the clinical model. In addition to these, DVH based metrics, specifically MLD and V20, and V5 were also included in the set of clinical predictors. Due to missing data for

significant number of the patients, the DVH-based metrics were computed manually using the tumor subtracted pulmonary volumes and the dose distributions. Table **??** lists all of the predictors and the variable type with which they were encoded in the analysis.

V_d of the lung receiving dose d given by:

$$V_{\mathrm{d}} = \frac{N_{\mathrm{Dose} \geq d}}{N_{\text{all voxels}}} \tag{2.1}$$

Bilateral MLD for each patient was also computed according to the following formula:

$$MLD = \frac{\sum^{eachvoxel} Dose}{N_{\text{all voxels}}} \tag{2.2}$$

2.7 Statistical Analysis

The differences between the patients characteristics (Table B.4 by RT technique (3DCRT vs IMRT) was assessed using a Wilcoxon test for continuous variables and Fisher Exact Test for categorical variables. The primary endpoint of the study was grade ≥ 2 RP per NCI CTCAE version 4.03. Both the univariate and multivariate analysis were stratified by radiation modality recognizing that the dosiomics feature performance may be a function of the type of dose-distribution a patient receives.

The predictive power of the top five features for each dataset was first assessed univariately using concordance index for binary response which is also equivalent to the area under the curve (AUC). Statistical significance is indicated for p-value ≤ 0.05 after accounting for multiple testing using the false-discovery-rate procedure by Benjamini and Hochberg [3].

A wide range of machine learning algorithms are available as packages in R. Choosing a specific model can be understood as a tradeoff between interpretibility and flexibility [31]. Each model has its own set of assumptions and bias. Models that fall in generalized linear models category, for example, Logistic Regression, tend to be more interpretable as they generally probe the additive effect of variables/features on prediction. This comes with a certain assumption about the underlying function that describes the phenomenon one is trying to model. Other non-parametric models provide more flexibility. For our multivariate analysis, we chose a simpler, more interpretble logistic regression for classification model as well as a less interpretable but flexible random forest classifier which is known for its ability to minimize over-fitting while at the same look inherent interaction effects between the

predictors (CT and Dose).

The logistic classification model is a member of the generalized linear model family. These algorithms are more intuitive to interpret and usually have no or very few hyper-parameters that require tuning during the training process. However, in their simplest implementation, only additive effect of variables can be probed.

Secondly, we implemented a random forest model. The RF algorithm is known to minimize over-fitting as well as implicitly capture interaction effects between the predictors. The analysis was implemented in R Statistical language [48] using the caret package [35]. Analysis workflow is depicted in Figure **??**. The data was iterated over 100 random data splits into training (70%) and test (30%) to assess the extent of predictive power. Models were internally validated using a 10-fold cross-validation. The median test AUC as well the range for each iteration are reported. Combined feature sets were derived from the mRMR selected radiomics and dosiomics features. In total, 7 feature sets or models were considered in our multivariate analysis - *a) Radiomics b) Dosiomics c) Radiomics+Dosiomics d) Clinical e) Radiomics+Dosiomics+Clinical*

The comparative performance of the model was only evaluated for the random forest models using a non-parametric permutation test based on the procedure described in [11]. Pairwise greater-than alternative hypotheses were investigated for combined models namely *Radiomics+Dosiomics* and *Radiomics+Dosiomics+Clinical* against the *Clinical* model. The ground truth binary labels were randomly resampled (k = 100) and a new RF model fitted. For each permutation, W_k statistic was extracted for a one-sided Wilcoxon test and compared with true label W_0. P-values were defined as follows:

$$p = \frac{1}{n_{\text{permut}} + 1} \sum_{k=1}^{n_{\text{permut}}} \begin{cases} \text{if } W_k > W_0 = 1 \\ \text{if } W_k < W_0 = 0 \end{cases} \tag{2.3}$$

2.8 Class Imbalance

Since machine learning algorithms are optimized to minimize the overall prediction error rate during training, special treatment is in order when analyzing datasets that are imbalanced. In our dataset, nearly 12% of patients developed RP consistent with the published prevalance rates. Due to this imbalanced binary response classes, machine learning models tend to learn the non-events better the events. Such models have high specificity but poor sensitivity.

There are established techniques available in order to address the issue of class imbalance.

Synthetic Minority Oversampling Technique (SMOTE) is a widely used technique to address class imbalance which uses K-Nearest Neigbor apporach to create new minority observations during the training process such that model learns both RP event cases and non-event cases equally [8]. The best model from training is then evaluated for performance on the hold-out set which still has the original ratio of event and non-events. SMOTE has been successfully implemented in various medical science application such as detecting breast cancer applications. The use of such resampling method is implemented within the cross-validated loop such that the validation results are always evaluated on data that is not affect by resampling and has the original event/non-event ratio.

Chapter 3

Univariate Results

3.1 Clinical Characterstics

Table B.4 shows the clinical characteristics of the 596 patients included in the analysis. The mean age for our cohort was 65.0 years divided evenly between the two genders. Over 90% of the patients were either current or former smokers. Nearly 85% of the patients were treated with concurrent chemotherapy. Most patients presented with stage IIIA disease (55.7%) followed by stage IIIB (32.7%). 85% of the patients were treated with concurrent chemotherapy and about 20% also underwent prior induction chemotherapy.

Majority of the patients were treated with 3DCRT (75.8%) and the remaining 24.2% were treated with IMRT (majority static field IMRT, n=5 volumetric modulated arc therapy). Out of 73 cases in total that developed grade≥2 RP, more patients (64%) were treated with 3DCRT than IMRT (p = 0.02). 12.2% of all patients were diagnosed with grade≥2 RP but significant more of the IMRT cases were diagnosed with RP (p = .02, Fisher Exact Test). The bilateral and ipsilateral V20 and V5 volumes (Table B.5) for the IMRT subgroup were significantly larger (p_{V20} = 0.04 , p_{V5} = 6.1e-4, Wilcoxon) compared to the patients treated with 3DCRT. The average MLD for the pooled cohort was 13.9 Gy, significantly different between the two modalities (p = 1.1e-6).

3.2 Clinical Predictors of Radiation Pneumonitis

Figure A-5a shows the univariate analysis for the pooled dataset. MLD was the strongest clinical predictor of grade ≥ RP (AUC = 0.70) followed by V20 (AUC = 0.70) and V5 (AUC = 0.65). While MLD and V20 were predictive in the pooled as well as the stratified datasets (Figure A-5b,c), the absolute performance of both predictors was noticeably higher for the IMRT subsample (AUC_{MLD} = 0.79, AUC_{V20} = 0.81) in contrast to the 3DCRT subsample (AUC_{MLD} = 0.64, AUC_{V20} = 0.62). Such an improvement was not seen for V5 which was statistically significant in predicting RP in all the datasets. Smoking was moderately positively associated with RP and the best predictor for the 3DCRT subgroup (AUC = 0.66). The V5 metric was found to be moderately predictive of RP in both 3DCRT (AUC = 0.60) and IMRT (AUC = 0.69) subgroup and statistically significant. Most of the patient-based metrics except age and smoking were not significantly associated with RP.

3.3 Bilateral Features

ROI-V20

Radiomics features extracted from the ROI-V20 were comparable in performance to ROI-V5 peritumoral features with textural complexity measures such as *wavelet.LHL_glcm_Imc1* and *wavelet.LHL_glcm_Imc2* showing moderate associations for the pooled and 3DCRT datasets. All five V20 radiomics features selected on the 3DCRT stratum were protective in nature and the best predictor was *wavelet.LHL_glcm_Imc2* (AUC = 0.65), consistent with the corresponding ROI-V5 3DCRT results. Additionally, *wavelet2.HHL_gldm_DependenceEntropy* (AUC = 0.65) and *wavelet2.LHH_glcm_Correlation* (AUC = 0.64) were also predictive and significant.

Radiomics features for the IMRT subsample showed relatively stronger performance overall. Most features were directly correlated with the probability of developing RP. The strongest predictors were *wavelet.LHH_gldm_DependenceNonUniformity* (AUC = 0.73) and *wavelet2.LHL_glszm_SmallAreaLowGrayLevelEmphasis* (AUC = 0.72).

The higher performance of dosiomics features for the IMRT subsample was also observed for the ROI-V20 features. *3D_firstorder_Maximum* dosiomics feature was significantly associated with RP all three datasets but the predictive power was highest for IMRT (AUC = 0.75). *wavelet2.LLH_glszm_GrayLevelNonUniformity* (AUC = 0.73) and *original_glszm_SizeZoneNonUniformity* (AUC = 0.71) were also significantly associated with RP for the IMRT cohort.

ROI-V5

Radiomics features selected in the pooled cohort were descriptors of texture and shape. *wavelet2.LLH_glcm_Imc2* was the strongest radiomics predictor and inversely correlated with RP with AUC = 0.63 and was statistically significant ($p<0.05$). Dosiomics features selected on the pooled dataset showed only slight associations with the AUC values ranging from 0.60 to 0.62.

Wavelet.LHL_glcm_Imc2 was selected as the best radiomics predictor for the 3DCRT subsample. Shape based features, *original_shape_Flatness* and *original_shape_Elongation* were the strongest radiomics features for the IMRT subsample with AUC=0.71 and AUC=0.68, respectively. The predictive performance of the dosiomics features for the IMRT subsample

ranged from 0.65 to 0.72 and *original_glszm_SizeZoneNonUniformity*, *Original_shape_Flatness* (AUC=0.71), *wavelet2.LHL_firstorder_Mean* (AUC=0.70) were strongest predictors. The 3DCRT selected dosiomics features were descriptors of the textural complexity but the univariate performance of these features, however, was not statistically significant.

3.4 Ipsilateral Features

ROI-V20

Radiomics features extracted from the ipsilateral ROI-V20 showed moderate univariate predictive power for RP predictions with AUC values ranging from 0. 0.61 - 0.65. Textural complexity descriptor *wavelet.LHL_glcm_Imc1* was once again the best radiomcis predictor (AUC = 0.65) for the pooled dataset followed by the ROI shape descriptor *original_shape_Maximum2DDiameterColumn* (AUC = 0.64).

The ipsilateral ROI-V20 dosiomics performance for the pooled and 3DCRT datasets was moderate. For the pooled dataset, the univariate performance ranged between 0.60-0.66 with *wavelet.HLL_glcm_Idmn* (0.66) and intensity-based feature, *log.sigma.5.0.mm.3D_firstorder_Maximum* (0.64) were the strongest dosiomics features for the pooled dataset.

When stratified by radiation technique, the mrmr radiomics and dosiomics features selected for the 3DCRT were quite similar. *wavelet.LHL_glcm_Imc2* was the best radiomcis predictor and *wavelet.HLL_glcm_Idmn* was the best dosiomics predictor for the 3DCRT dataset. While the performance for the 3DCRT dataset was modest, the all the features associated with RP were found to statistically significant.

Most of the top 5 mrmr radiomics and dosiomics features selected on the IMRT subgroup showed particularly stronger prognostication power to predict RP. Textural radiomics feature, *wavelet.LHH_gldm_DependenceNonUniformity* (AUC = 0.74) and filtered first order dosiomics feature, *log.sigma.5.0.mm.3D_firstorder_Maximum* (AUC = 0.74) exhibited the strongest associations. Dosiomics dose coarseness descriptor, *wavelet.LHL_ngtdm_Coarseness* (AUC = 0.71) and radiomics coarseness predictor *log.sigma.2.0.mm.3D_ngtdm_Coarseness* (AUC = 0.70) were both negatively associated with RP.

ROI-V5

Ipsilateral ROI-V5 Radiomics and dosiomcis features selected in the pooled cohort were mostly descriptors of texture. *wavelet2.LLH_glcm_Imc2* was the strongest radiomics predictor and inversely correlated with RP with AUC = 0.63 and was statistically significant ($p<0.05$). Most of the dosiomics features other than the strongest dosiomics predictor, *wavelet.HLL_ngtdm_Contrast* ( ) showed poor predictive power for the pooled dataset, a result that also carried over into the dosiomics performance when the datasets were stratified.

The 3DCRT dosiomics features were not good predictors of RP with only two of the five selected dosiomics features showing statistically significant predictive power. In contrast, the IMRT subgroup univariate performance was stronger and the radiomics and dosiomics features selected were dissimilar to the ones selected in the pooled or 3DCRT datasets. Texture non-uniformity radiomics and dosiomics statistics were strongly associated with RP. Overall, the performance of features selected on the IMRT subgroup was around 0.70 or more in the univariate setting. Moreover, all of radiomics and dosiomics features were statistically significant based on the criteria described in Section 2.7.

Chapter 4

Logistic Classification Models

4.1 Bilateral Models

ROI-V20

The logistic classifier built using bilateral features from ROI-V20 (Figure A-14) exhibited poor performance for the pooled dataset. The model performance ranged from 0.64 - 0.68 (median AUC). The *Clinical* model had the highest median AUC (0.68) and the radiomics model was the lowest performing model.

Radiomics+Dosiomics performance (AUC = 0.70) was appreciably higher for the 3DCRT subgroup than the *Clinical* model (AUC = 0.64) (Figure A-14). The *Radiomics+Dosiomics* model was also the best performing model for the 3DCRT subgroup. Including the clinical predictors into the *Radiomics+Dosiomics* model did result in a modest improvement in the predictive power compared to the *Clinical* model. The dosiomics features used alone were a poor predictor of RP (AUC=0.60).

In contrast, the *Dosiomics* model performance for the IMRT subgroup (Figure A-14), was significantly stronger (AUC = 0.73). The *Clinical* model was the best performing model (AUC = 0.77) and adding the *Radiomics+Dosiomics* predictors with the *Clinical* model did not reveal an enhancement in predictive power.

ROI-V5

Generally, the logistic classifier built using bilateral features from ROI-V5 (Figure A-15) did not vary vastly in performance for the pooled dataset when compared to the ROI-V20 logistic models. The median AUC had a particularly narrow range (0.67-0.70) for the pooled dataset, with *Radiomics+Dosiomics+Clinical* noteably performing the strongest.

The models fit on datasets stratified by radiation modality revealed some difference, especially for the 3DCRT subgroup when compared with the ROI-V20 bilateral logistic model. The model performance for the 3DCRT subgroup did not show a improved *Dosiomics* or *Radiomics+Dosiomics* performance. In fact, the 3DCRT model relative performance was similar to the pooled dataset and an overall decrease the predictive performance was observed (Figure A-15). The models again had a very narrow ranged with median AUC ranging from 0.64-0.65.

The *Dosiomics* model (AUC =0.78) was the strongest performing model (Figure A-15). However, including the combining the radiomics and dosiomics features in the *Ra-*

diomics+Dosiomics model did not further improve the performance in logistic regression setting. The *Clinical* model (AUC=0.77) performed comparably to the *Dosiomics* model. Combining the *Radiomics+Dosiomics* and *Clinical* features together did not exhibit an increase in the predictive power compared to the *Clinical* model.

4.2 Ipsilateral Model

ROI-V20

The logistic classifier built using ipsilateral ROI-V20 features (Figure A-14) exhibited modest performance across the pooled, 3DCRT and IMRT datasets.

The pooled *Radiomics+Dosiomics* (AUC = 0.67) model performance was comparable to the *Clinical* (AUC = 0.67). The *Radiomics+Dosiomics+Clinical* model (AUC = 0.69) did show slight improvement over the *Clinical* model.

The *Radiomics+Dosiomics* model (AUC = 0.70) was the again the best performing model out of the classifiers fit on the 3DCRT subgroup. Combining the *Radiomics+Dosiomics* predictors with the *Clinical* (AUC = 0.66) did not enhance the predictive performance of the *Radiomics+Dosiomics+Clinical* (AUC =0.66) model.

IMRT subgroup models showed peculiar change in performance in that combining the *Radiomics* (AUC = 0.69) and *Dosiomics* (AUC = 0.71) features together weaker model (*Radiomics+Dosiomics* AUC = 0.64) compared to the constituent models for the ipsilateral ROI-V20 features. The *Clinical* model (AUC = 0.77) was the best performing model.

ROI-V5

Radiomics+Dosiomics+Clinical model (AUC = 0.71) for the pooled dataset showed an appreciable increase in predictive power compared to the ROI-V20 ipsilateral *Radiomics+Dosiomics+Clinical* (AUC=0.69) model as well as the corresponding *Clinical* model (AUC = 0.67) (Figure A-17).

For the 3DCRT subgroup, all models showed comparable modest performance with *Dosiomics* showing the weakest performance (AUC = 0.65) and the *Radiomics* (AUC = 0.69) model showing the strongest performance followed by the *Radiomics+Dosiomics* model (AUC = 0.68) (Figure A-17.

The *Clinical* model was the best performing model for the IMRT subgroup far outperforming the individual or composite models. Interestingly, similar to ROI-V20 ipsilateral

models, the composite *Radiomics+Dosiomics* model showed a decrease in performance in logistic setting.

Chapter 5

Random Forest Classification Models

5.1 Bilateral Models

ROI-V20

Figure A-10 shows the test AUC plotted for the bilateral ROI-V20 RF models stratified by radiation modality. For the pooled dataset (Figure A-10, left), the *Dosiomics* model (AUC = 0.66) had a stronger performance than the *Radiomics* model (AUC = 0.60) and the difference was statistically significant (p = 0.01). The combined *Radiomics+Dosiomics* model (AUC = 0.71) improved significantly over the individual models. The *Clinical* model was not found to be statistically better than the *Radiomics+Dosiomics* model and vice-versa. The combined *Radiomics+Dosiomics+Clinical* (AUC = 0.76) model was the best performing model and statistically significantly better than the *Clinical* and *Radiomics+Dosiomics* model (p<0.01).

Radiomics+Dosiomics performance (AUC = 0.76) was appreciably higher for 3DCRT (Figure A-10, middle) and statistically better (p = 0.01) than the *Clinical* model (AUC = 0.60) (Figure **??**a). Including the clinical predictors into the *Radiomics+Dosiomics* model did not result in an improvement in the predictive power. The *Radiomics+Dosiomics+Clinical* model was statistically better (p<0.01) than the *Clinical* model.

For the IMRT subgroup (Figure A-10, right), contrary to ROI-V5, an overall stronger trend across all models was not observed. Notwithstanding, the *Radiomics+Dosiomics* model (AUC = 0.72) was still statistically stronger than the *Radiomics* model (AUC = 0.67, p= 0.07). *Radiomics+Dosiomics* and *Clinical* model were at par with each other. Including the *Radiomics+Dosiomics* predictors to the *Clinical* model, significantly enhanced the predictive power of the *Clinical* model (AUC = 0.77).

ROI-V5

Figure A-11 shows the test AUC plotted for the bilateral ROI-V5 RF models stratified by radiation modality. For the pooled analysis (Figure A-11, left), while the individual *Radiomics* (AUC = 0.59) and *Dosiomics* (AUC = 0.59) model performance was poor, *Radiomics+Dosiomics* model (AUC = 0.65) showed significant (p <0.05) improvement in predictive power over the respective individual models. *Radiomics+Dosiomics+Clinical* model performed the best (AUC = 0.73) in the pooled analysis but was not statistically better than the *Clinical* model.

Stratifying by radiation modality revealed our approach's true benefit. For the 3DCRT

subgroup ((Figure A-11, middle), model performance was mostly comparable to the pooled cohort, except that the combined *Radiomics+Dosiomics+Clinical* model was statistically better ($p < 0.01$) than the *Clinical* model. Models derived from radiomics and dosiomics features exhibited stronger predictive performance for IMRT (Figure A-11, right). The *Radiomics* model (AUC = 0.72) showed moderate performance as did the *Dosiomics* model (AUC = 0.71). The *Radiomics+Dosiomics* model was statistically better than the individual models ($p \leq 0.01$). Statistically significant performance gains were again observed for the combined *Radiomics+Dosiomics* model (AUC = 0.77) and *Radiomics+Dosiomics+Clinical* model (AUC = 0.81) relative to the *Clinical* model (AUC = 0.73).

5.2 Ipsilateral Model

ROI-V20

Figure A-12 shows the test AUC plotted for the ipsilateral ROI-V20 RF models stratified by radiation modality. The *Dosiomics* predictors were more effective at predicting RP than the *Radiomics* model (Figure A-12, left). An increase in the overall AUC was observed for the *Radiomics+Dosiomics* model but the *Clinical* (AUC = 0.73) model was statistically stronger (p *le*0.01) than the *Radiomics+Dosiomics* (AUC = 0.68) model. Including the *Radiomics+Dosiomics* model with the *Clinical* mode, *Radiomics+Dosiomics+Clinical* (AUC = 0.75), did slightly (p = 0.06) improve the model over the *Clinical* model alone but strictly speaking, the difference was not statistically significant according the p *le* 0.05 criteria used for our study.

The performance of the ROI-V20 ipsilateral radiomics and dosiomics features exhibited particularly stronger combined performance for the 3DCRT subgroup (Figure A-12, middle). Specifically, the *Radiomics+Dosiomics* alone was statistically stronger (p *le* 0.05) than the *Clinical* (AUC = 0.69) model. Including the *Radiomics+Dosiomics* predictors with the *Clinical* further improved the RP predictive performance (*Radiomics+Dosiomics+Clinical*, AUC = 0.77).

For the IMRT subgroup, the *Radiomics* and *Dosiomics* fared comparably (AUC = 0.68) ((Figure A-12, right)). The combined performance of the features (*Radiomics+Dosiomics*, AUC = 0.75)) was better than its individual constituents and based on the permutation test results, the *Clinical* was not statistically better than the *Radiomics+Dosiomics* model.

The *Radiomics+Dosiomics+Clinical* (AUC = 0.77) model, however, was the best performing model and enhancement over the *Clinical* model alone was statistically significant (p *le* 0.05).

ROI-V5

Figure A-13 shows the test AUC plotted for the ipsilateral ROI-V5 RF models stratified by radiation modality. The pooled performance (Figure A-13, left) of the individual ROI-V5 ipsilateral *Dosiomics* was slightly better than random (AUC = 0.59). The *Clinical* (AUC = 0.73) model was able to outperform (p *le* 0.01) the *Radiomics+Dosiomics* (AUC = 0.68) model. Including the combined feature set with the clinical predictors augmented the predictive ability of the model (*Radiomics+Dosiomics+Clinical*, AUC = 0.75). The difference between the *Radiomics+Dosiomics+Clinical* and *Clinical* model was statistically significant (p *le* 0.01).

The 3DCRT subgroup *Radiomics+Dosiomics* model was once again the second best model (Figure A-13, middle), far outperforming its individual constituent *Radiomics* and *Dosiomics* model as well as the *Clinical* model. The difference between the *Radiomics+Dosiomics* amd *Clinical* model was statistically significant. As expected, combining the clinical predictors with the *Radiomics+Dosiomics* model, improved test AUC slightly but the overall model performance was still statistically stronger than the *Clinical* model.

Chapter 6

Discussion

As a result of recent advances in radiation delivery modalities, thoracic radio-oncologists have successfully been able to deliver higher therapeutic doses of radiation in an attempt to affect better outcomes. Despite the superiority of IMRT, owing to higher dose conformality and normal tissue sparing, the locoregional failure rates remains fairly high as compared to other sites. Over 40% of the patients suffer from loco-regional recurrence event within 2 years of initial course of radiation (Figure A-18). Moreover, while IMRT has proven to be protective over conventional 3DCRT [10], RP is still a major consideration for the treatment of locally advanced-stage lung cancer patients. Achieving tighter local control requires safely escalating the prescription dose to the tumor but this necessitates assessing patient-specific pre-treatment risk to develop RP.

Our efforts described in this work centered around evaluating the predictive performance of radiomics and dosiomics features to predict RP in a stratified analysis. The ROI dependence of these features was assessed by using two types of ROI, namely ROI-V20 and ROI-V5. Furthermore, the lung-sided was also considered by computing radiomics and dosiomics features from the bilateral and ipsilateral lung.

The univariate performance of both radiomics and dosiomics features for the 3DCRT subgroup was moderate to poor. The bilateral mRMR radiomics features selected on the 3DCRT for ROI-V20 and ROI-V5 were protective in nature. This was also largely true for the ipsilateral features. Such a trend was not seen for the dosiomics features selected on the 3DCRT. Protective imaging biomarkers can help physicians in safely escalating the prescription for patients that are less likely to develop RP.

Compared to ROI-V5, ROI-V20 ipsilateral and bilateral dosiomics features peformance was stronger and all but one of the features were statistically significant. This was not true for the ROI-V5 dosiomics features. Our univariate analysis also showed that the radiomics and dosiomics features selected on the IMRT subgroup were comparatively much stronger predictors of RP compared to the 3DCRT features. In contrast to the 3DCRT, no strict directionality was observed for the bilateral and ipsilateral radiomics features.

Bilateral ROI-V5 shape-based features, such as Flatness and Elongation were stronger predictors of RP over other textural and intensity based features for the IMRT subgroup (Figure). These findings highlight that the spatial distribution of the V5 dose, in addition to the DVH-based metrics like V20 and MLD (Figure A-5, may be vitally important for accurately assessing RP risk for a particular patient. High-level features extracted from

dose-based ROIs are able to provide new information that is unused when a plan is evaluated using V20 or MLD alone. In addition, for IMRT patients HU textural non-uniformity and dose non-uniformity measures for the ipsilateral lung were positively correlate with RP (Figure **??**,A-8)

For the multivariate analysis, a logistic classification and the random forest classification models were trained and tested using the radiomics, dosiomics, radiomics+dosiomics, radiomics+dosiomics+clinical, and clinical feature sets. The logistic classifiers were on average performed better than the corresponding random forest models for the individual radiomics and dosiomics features (Figure A-14, A-16, A-15, A-17).

However, the logistic performance for the combined feature sets was underwhelming. However, random forest *Radiomics+Dosiomics* and *Radiomics+Dosiomics+Clinical* models were effective at predicting RP and statistically better than *Clinical* model. These findings confirm the assumption in our initial hypotheses that the interaction effects between the CT features and dose features may play an important role in predicting RP compared to the DVH-based clinical model.

We find that the radiodosiomics approach exhibits stronger performance for IMRT across all the models. The dosiomics features may be intrinsically better suited for IMRT-type dose distributions as they tend to be spatially more hetrogeneous than 3DCRT dose distributions. It could also explain in part why the radiomics and dosiomics features perform particularly well on IMRT cohort. Therefore, the effectiveness of our approach on a more general patient population must be proven before clinical adaptation.

The dose-based ROIs utilized for feature extraction manifest itself as performance variation between the two radiation modalities. Higher conformality and tighter dose-distributions in IMRT results in greater overlap with the pulmonary tissue which V5 is able to represent. Meanwhile, for 3DCRT V5 provides very little additional pulmonary overlap in addition to the ROI-V20. As such, the 3DCRT *Radiomics+Dosiomics* model generated using the ROI-V20 performed better compared to ROI-V5 ($p < 0.01$). The opposite was true for IMRT in that the ROI-V5 *Radiomics+Dosiomics* model was statistically stronger than the V20 counterpart ($p = 0.02$) as well as the *Clinical* model ($p < 0.01$). In other words, the role of ipsilateral V20 and features extracted from that region are strongly predictive of RP risk in patients receiving 3DCRT. On the other hand, the bilateral V5 features are strongly predictive of RP in patients receiving IMRT.

The aforementioned findings were also limited by numerous factors and these caveats are worth mentioning explicitly. Firstly, due to the unavailability of a external validation dataset, the random forest and logistic machine-learning models were validated internally using 10-fold cross-validation using 70%/30% train/test data partition. We believe that while using a 10-fold cv approach was sufficient for this study, the performance of our models must be evaluated on an external dataset before clinical adoption. Moreover, the training process was also limited by the low number of RP cases in the dataset. This was partly remedied in our analysis by employing SMOTE resampling technique. Our findings will need to be validated on a much larger dataset, even though we anticipate that any RP related dataset will likely have low event rate, consistent with the published RP prevalence rates for NSCLC treated with CRT, however, a model trained on larger number of RP cases will ensure a more robust RP assessment with lower AUC variability. Alternatively, a larger dataset may allow for a two-step training process in which events and non-event models are trained and test separately. The final binary RP predictions would then be pooled using boolean logic. Such an approach could help physicians in classifying cases that will benefit from dose-escalation.

The cohort analyzed in this study spans a time-frame which also coincides with the introduction of IMRT at BWH (Figure **??**. A high majority of the IMRT cases considered in our study were selected by the physicians to the receive the newly commissioned modality. This selection bias means that the IMRT cases in our dataset were inherently higher-risk as seen by the larger PTV volumes (Table B.5). The efficacy of the radiodosiomics approach must be evaluated on an IMRT/VMAT dataset that includes patients representing a wide crosssection of disease and risk.

The study of radiomics and dosiomics is based in features generated from certain ROI. While the ROI method is great for generating building blocks of new information, they tend to make the training process tedious and subjective. It will be interesting to use some of intuition developed in our work to implement a multi-modality or multi-input deep learning based approach that ties together the CT and Dose spatial information.

Chapter 7

Conclusion

We describe a promising new approach to model RP risk for patients being treated with radiation treatment for locally-advanced lung cancer that utilizes pre-existing CT image data and treatment plan dose-distributions. Our multivariate results prove that incoporating dosiomics features along with radiomics features in a random forest classifier allows for an improved prediction of RP, especially for IMRT-type dose distributions. While the combined radiomics and dosiomics approach performance was stronger for IMRT dataset, the *Radiomics+Dosiomics* model was extremely effective at predicting RP for 3DCRT and significantly ($p \leq 0.05$) better than the *Clinical* model. The goal of such an approach is to assist physicians in customizing a treatment prescription based on patient's risk level and type of prescribed radiation treatment. A patient with a low pre-treatment RP risk could be treated with a more aggressive radiation prescription in order achieve excellent locoregional control. Conversely, a patient with a high pre-treatment risk could be considered for more conservative dose prescription. In a sense, our approach is able to capture a patient's baseline differences in anatomy and dose distributions.

Appendix A

Figures

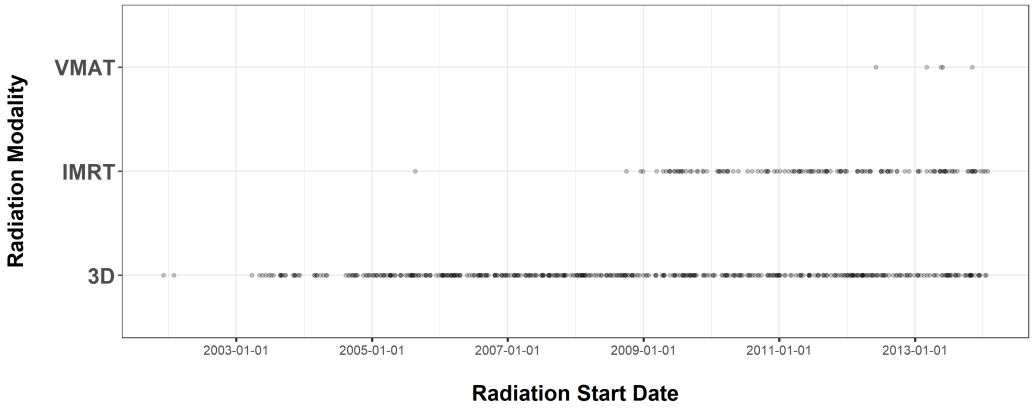

Figure A-1: Radiation Modality as a function treatment start date. Introduction of IMRT at BWH. High risk patients elevated to IMRT instead of 3DCRT initially

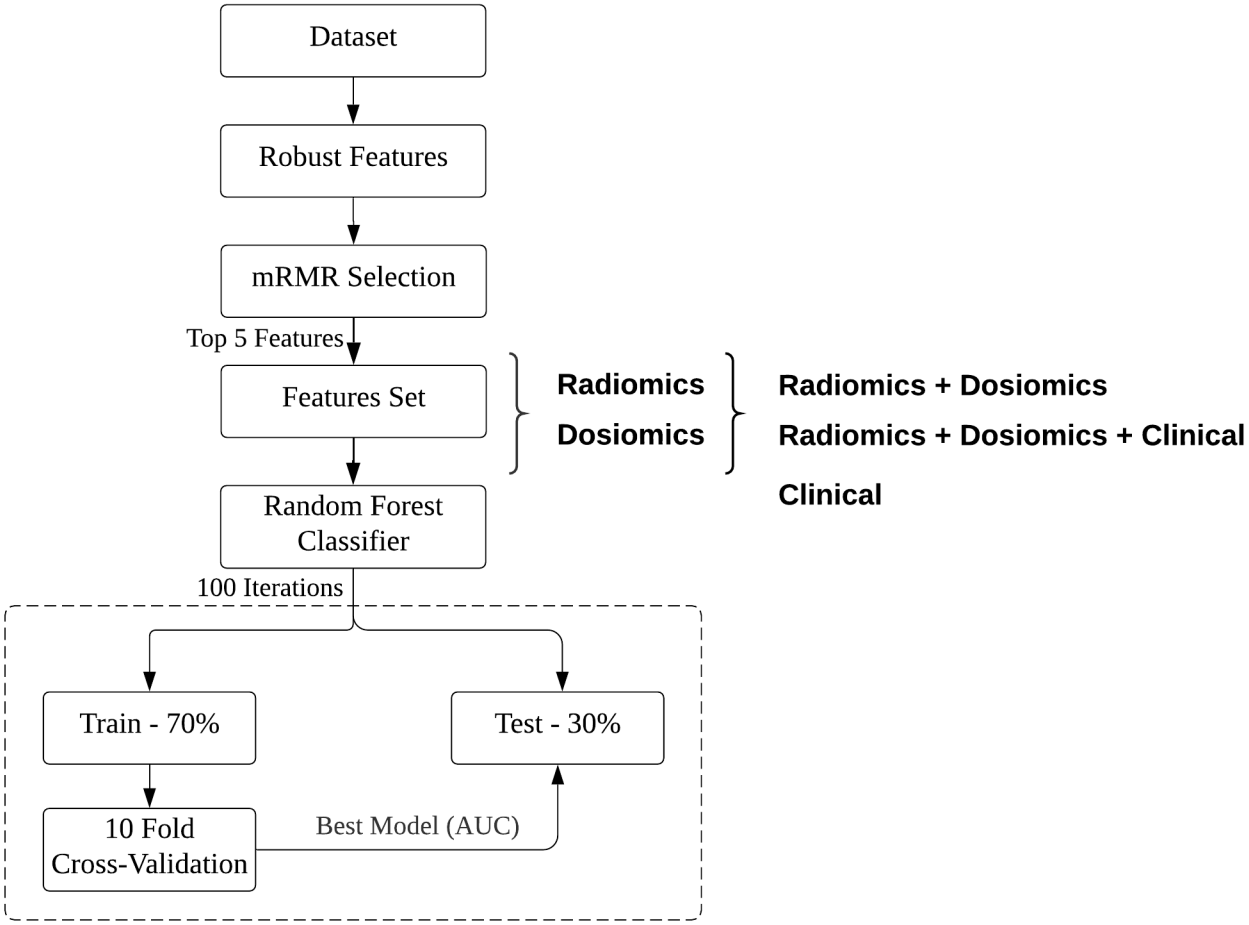

Figure A-2: Machine Learning Architecture for RP Prediction

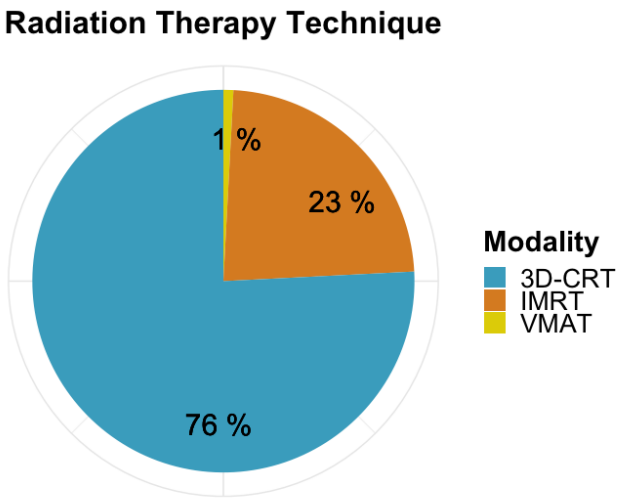

Figure A-3: Radiation Modality

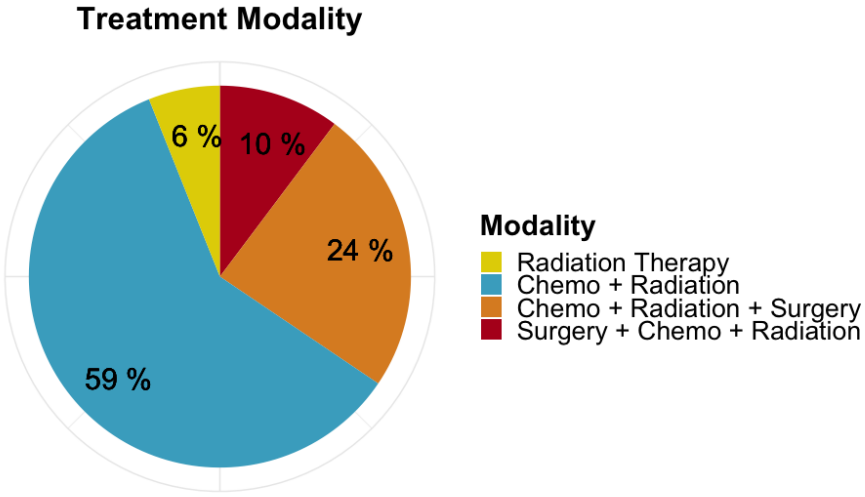

Figure A-4: Treatment Modality

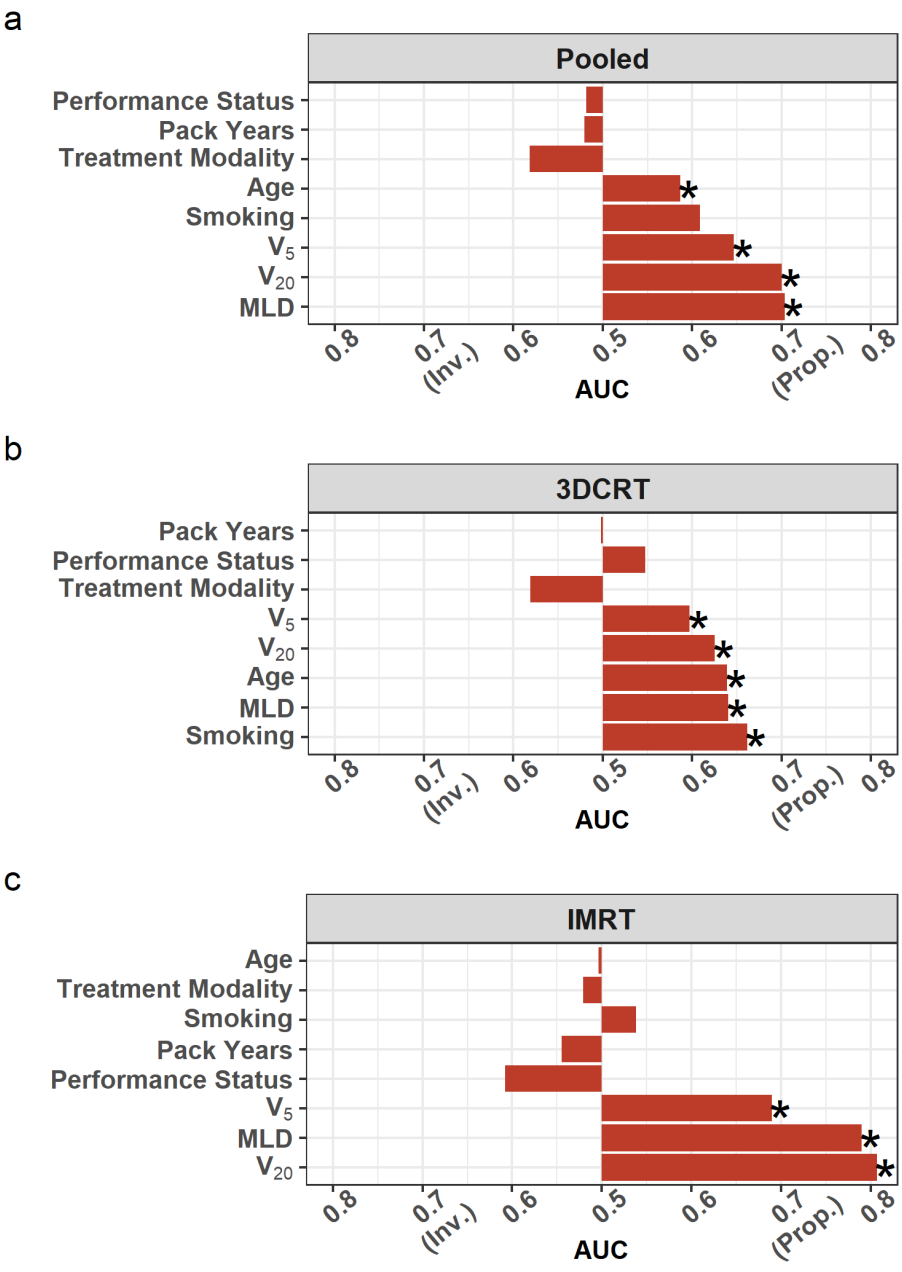

Figure A-5: Univariate predictive perform of clinical predictors for the pooled dataset and stratified by radiation modality

V_{20} Bilateral Univariate Performance

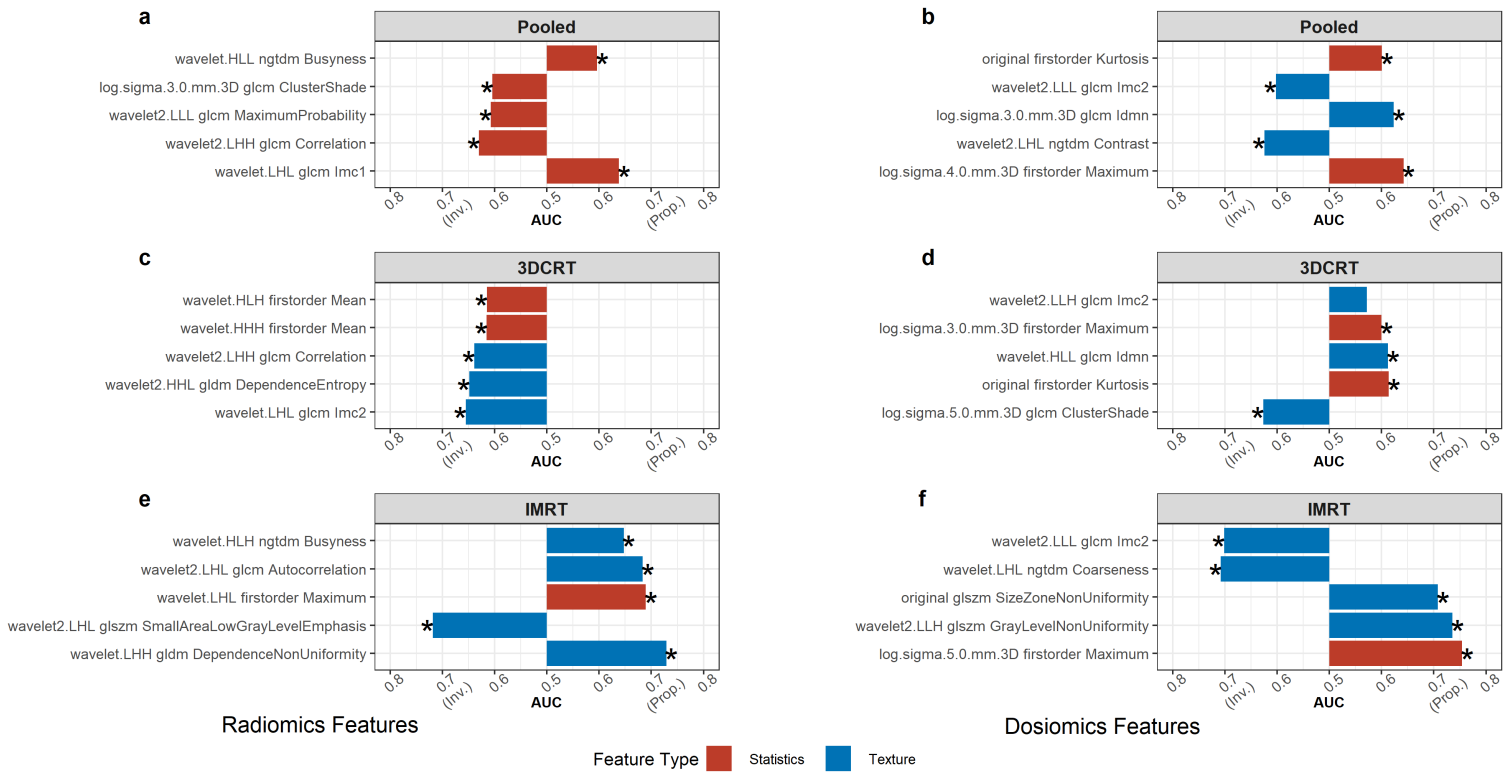

Figure A-6: Univariate predictive performance of the the top 5 mRMR selected bilateral ROI-V20 Radiomics (left) and Dosiomics (right) features

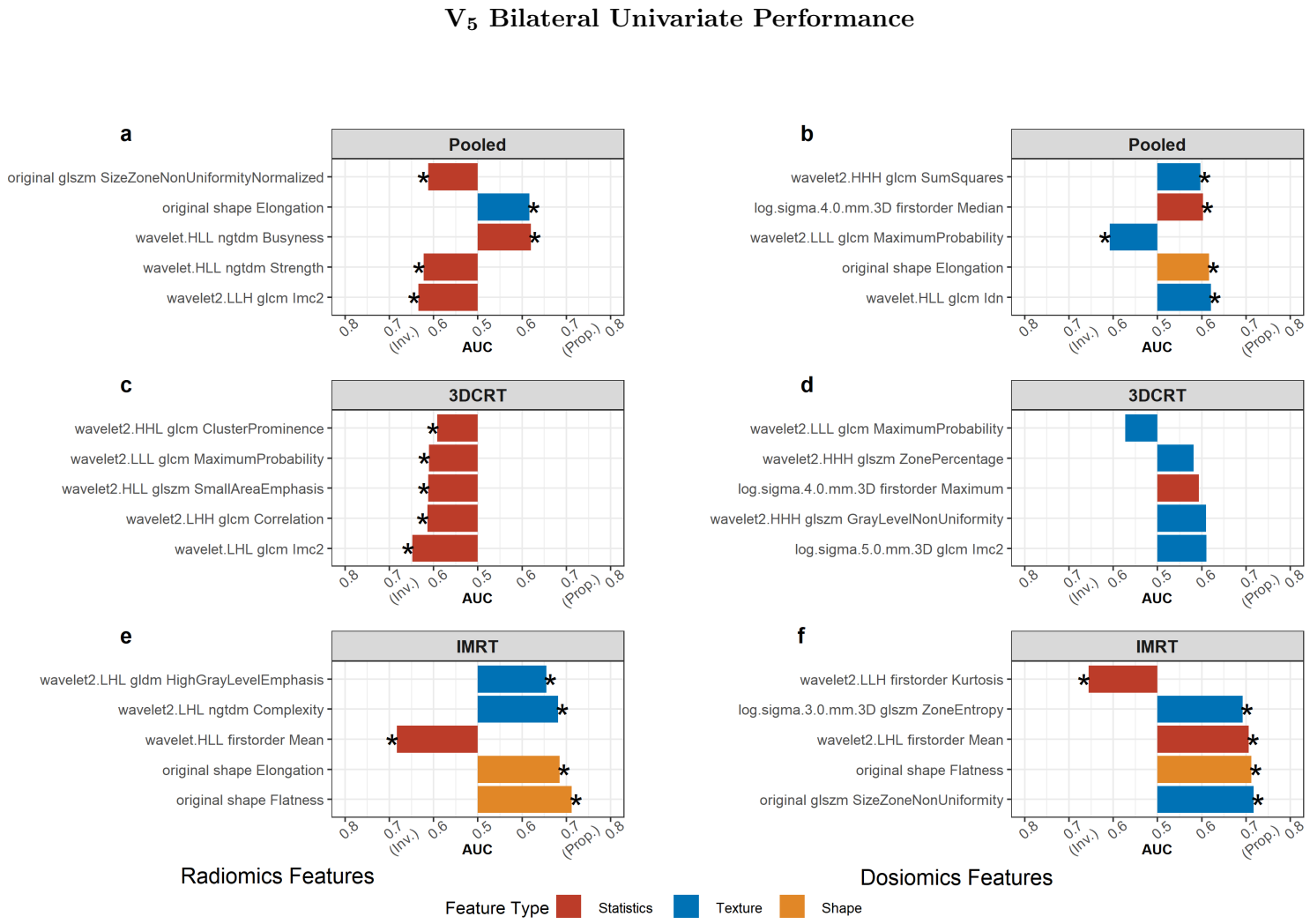

Figure A-7: Univariate predictive performance of the the top 5 mRMR selected bilateral ROI-V5 Radiomics (left) and Dosiomics (right) features

V_{20} Ipsilateral Univariate Performance

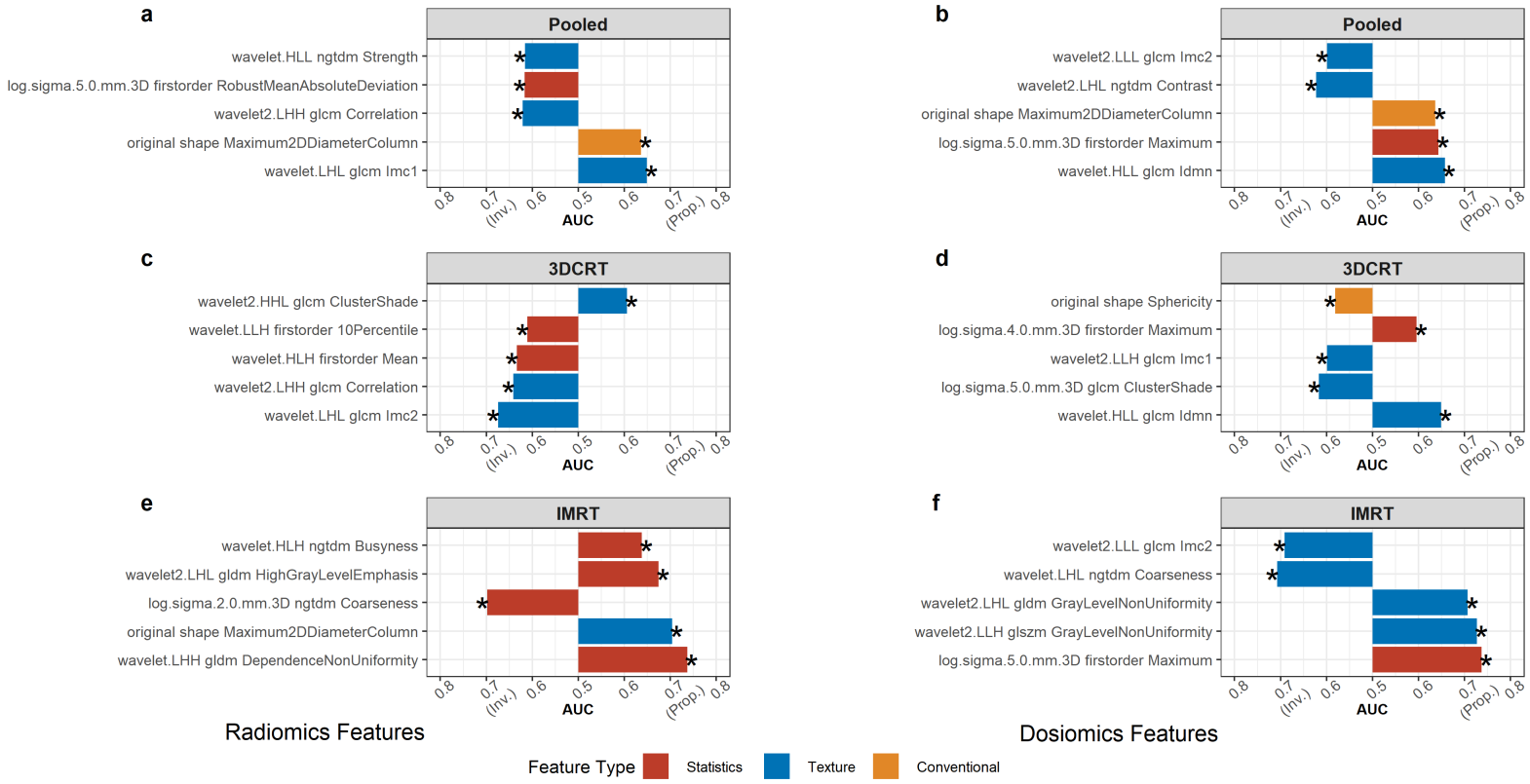

Figure A-8: Univariate predictive performance of the the top 5 mRMR selected ipsilateral ROI-V20 Radiomics (left) and Dosiomics (right) features

V_5 Ipsilateral Univariate Performance

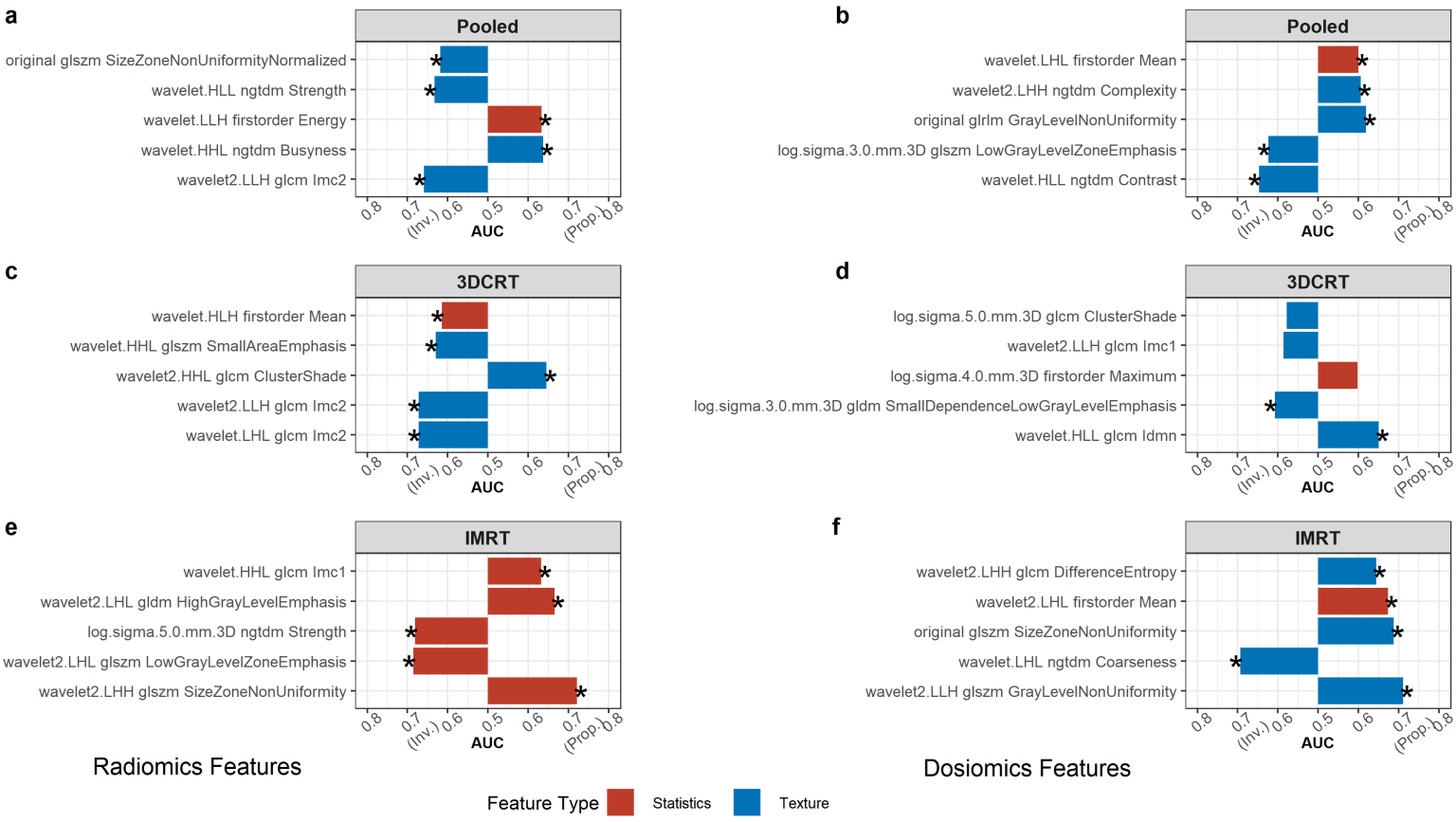

Figure A-9: Univariate predictive performance of the the top 5 mRMR selected ipsilateral ROI-V5 Radiomics (left) and Dosiomics (right) features

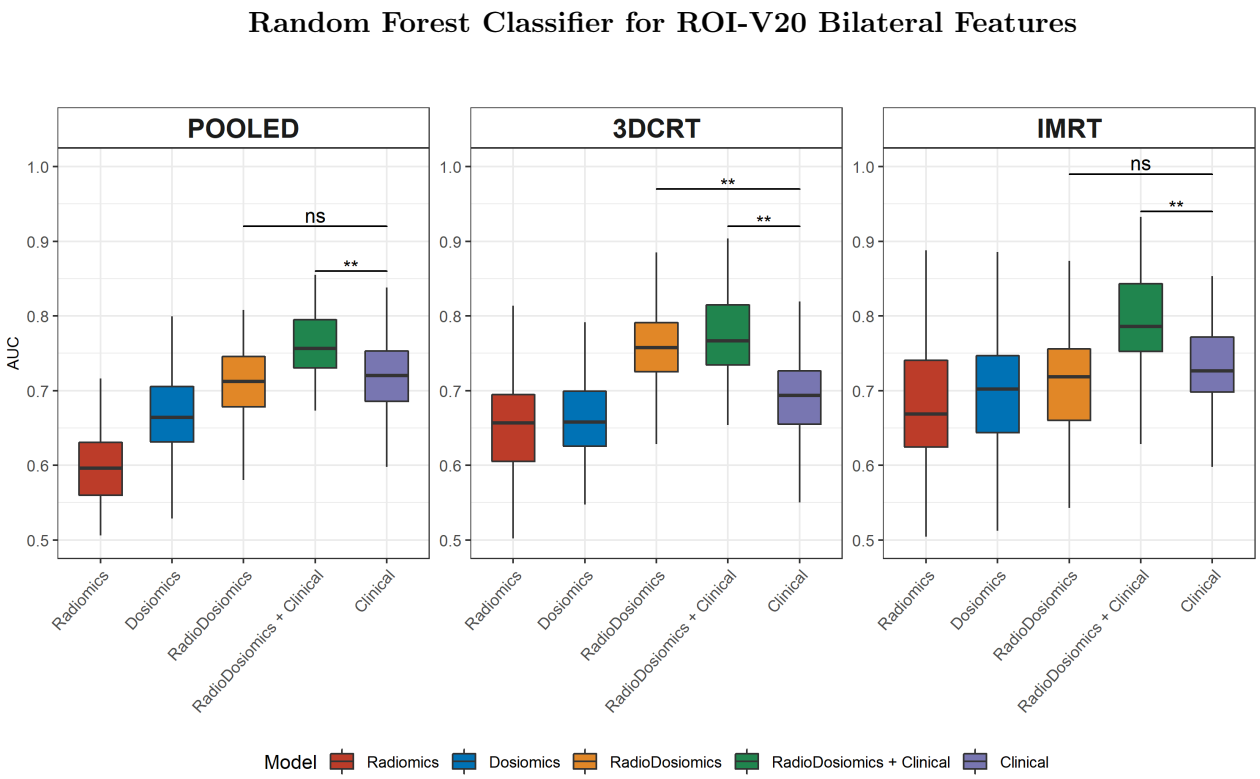

Figure A-10: Random Forest performance of bilateral ROI-V20 features compared against the clinical model using permutation test (** - p≤0.01, * - p≤0.05). Radiodosiomics model augments the clinical model but the gains are more noticeable in the IMRT subgroup

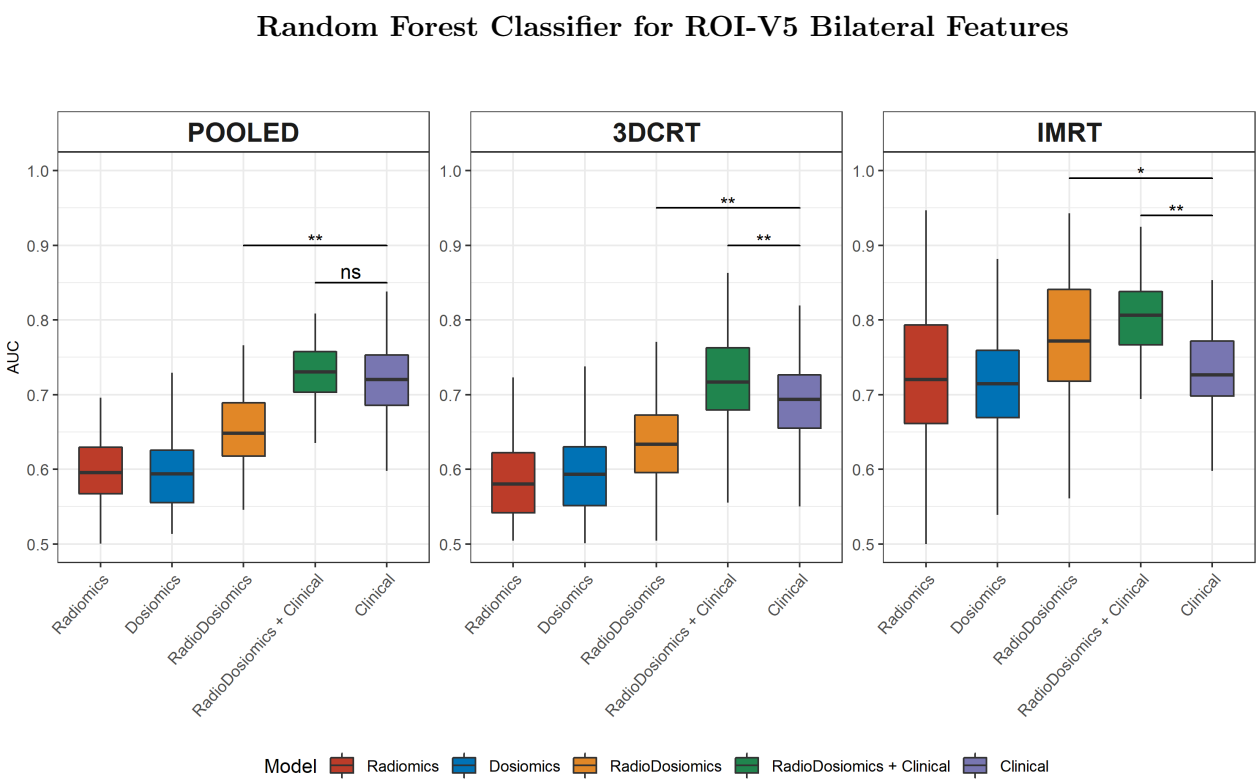

Figure A-11: Random Forest performance of bilateral ROI-V5 features compared against the clinical model using permutation test (** - p≤0.01, * - p≤0.05). Radiodosiomics (Radiomics+Dosiomics) model augments the clinical model. Absolute performance is especially stronger for the IMRT subsample

Random Forest Classifier for ROI-V20 Ipsilateral Features

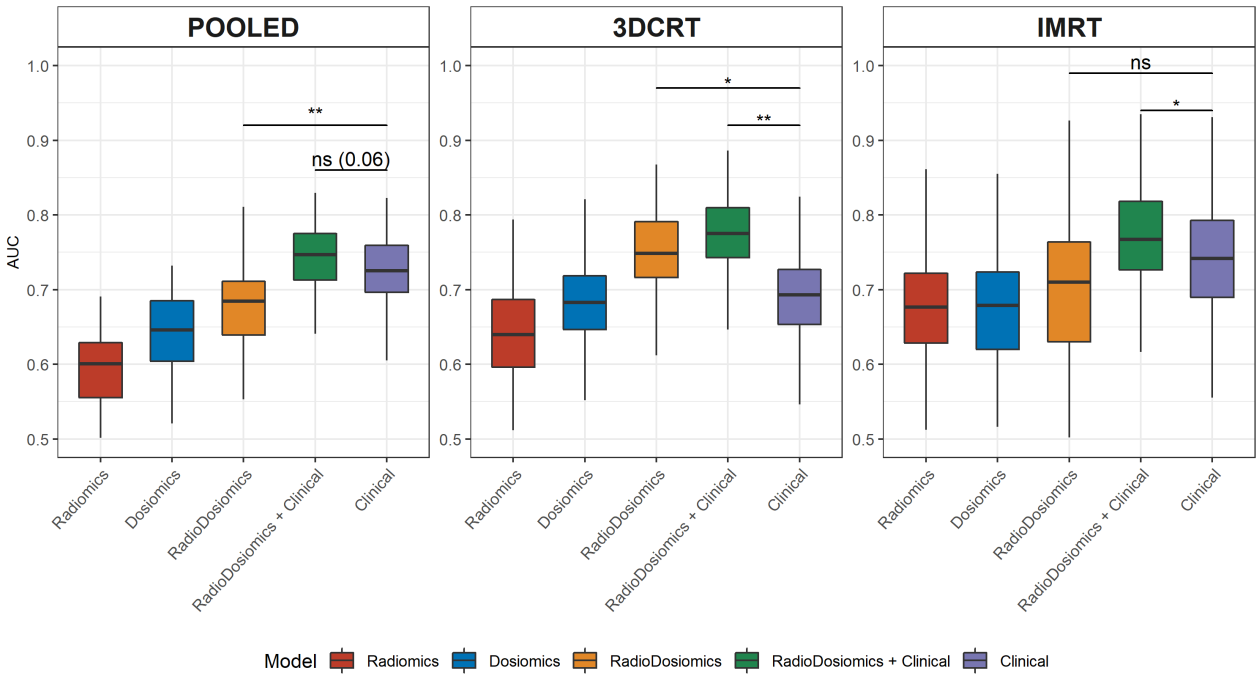

Figure A-12: Random Forest performance of ipsilateral ROI-V20 features compared against the clinical model using permutation test (** - p≤0.01, * - p≤0.05). Radiodosiomics model augments the clinical model but the gains are more noticeable in the IMRT subsample

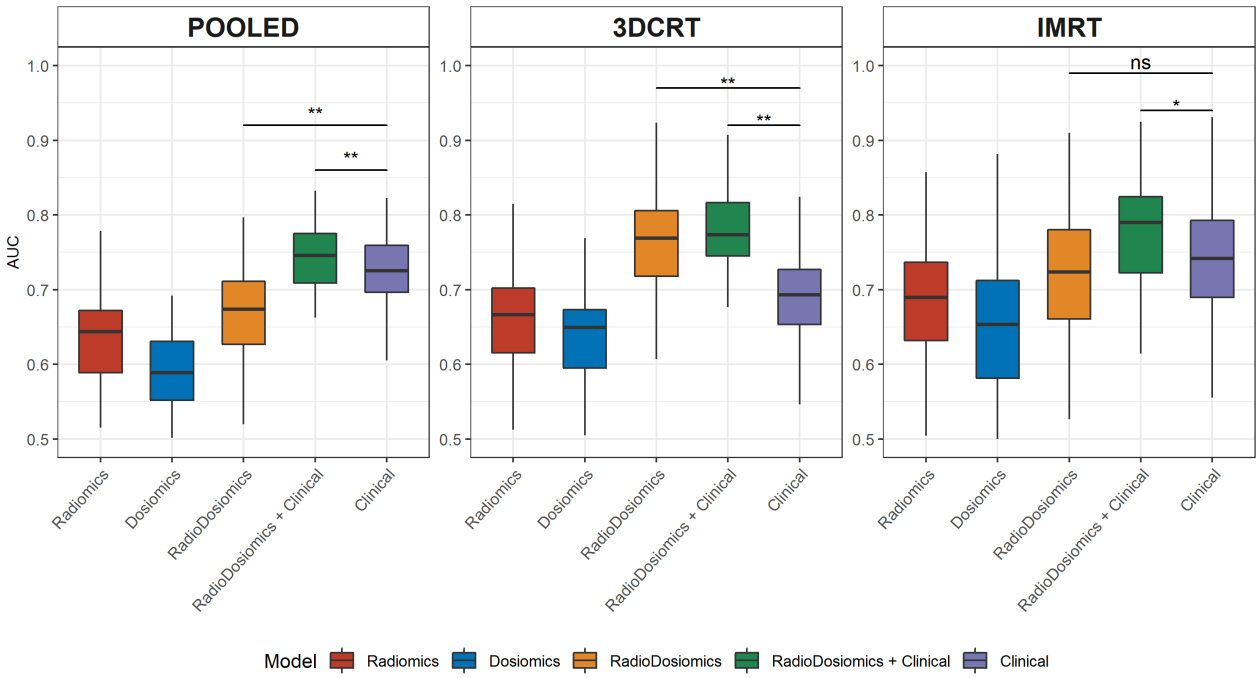

Figure A-13: Random Forest performance of ipsilateral ROI-V5 features compared against the clinical model using permutation test (** - p≤0.01, * - p≤0.05). Radiodosiomics (Radiomics+Dosiomics) model augments the clinical model. Absolute performance is especially stronger for the IMRT subsample

Logistic Classifier for ROI-V20 Bilateral Features

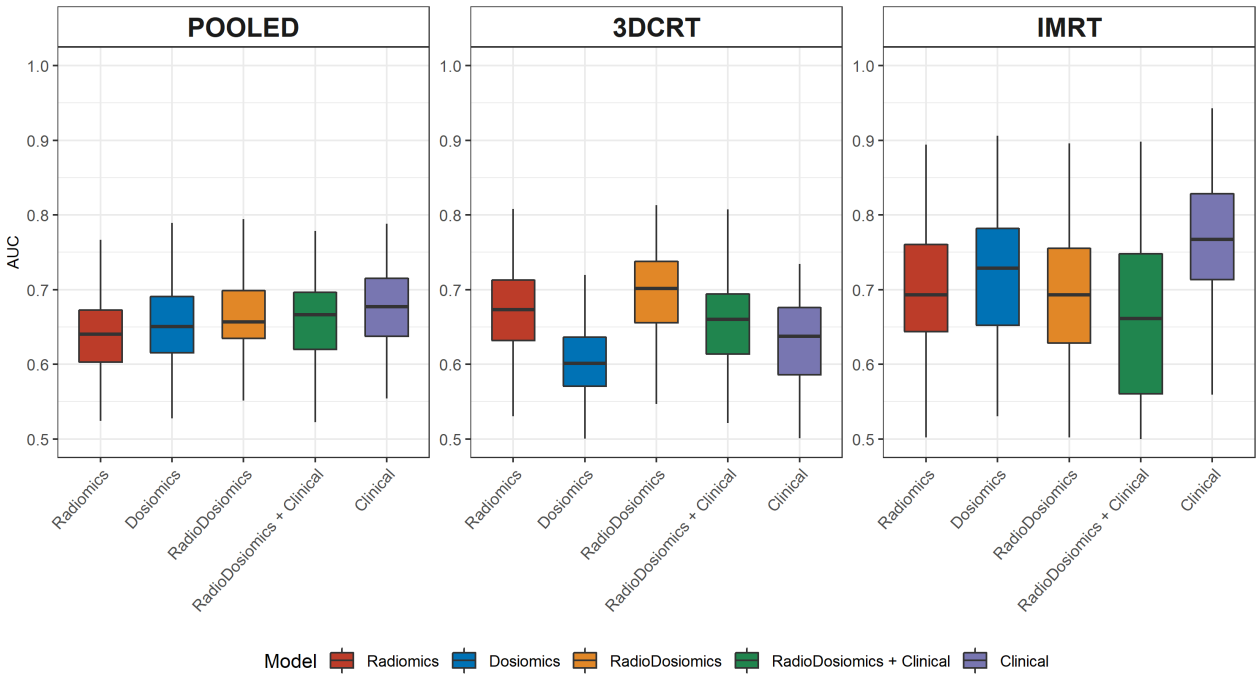

Figure A-14: Logistic classifier performance of the bilateral ROI-V20 radiomics, dosiomics and the combined models plotted along with the clinical model

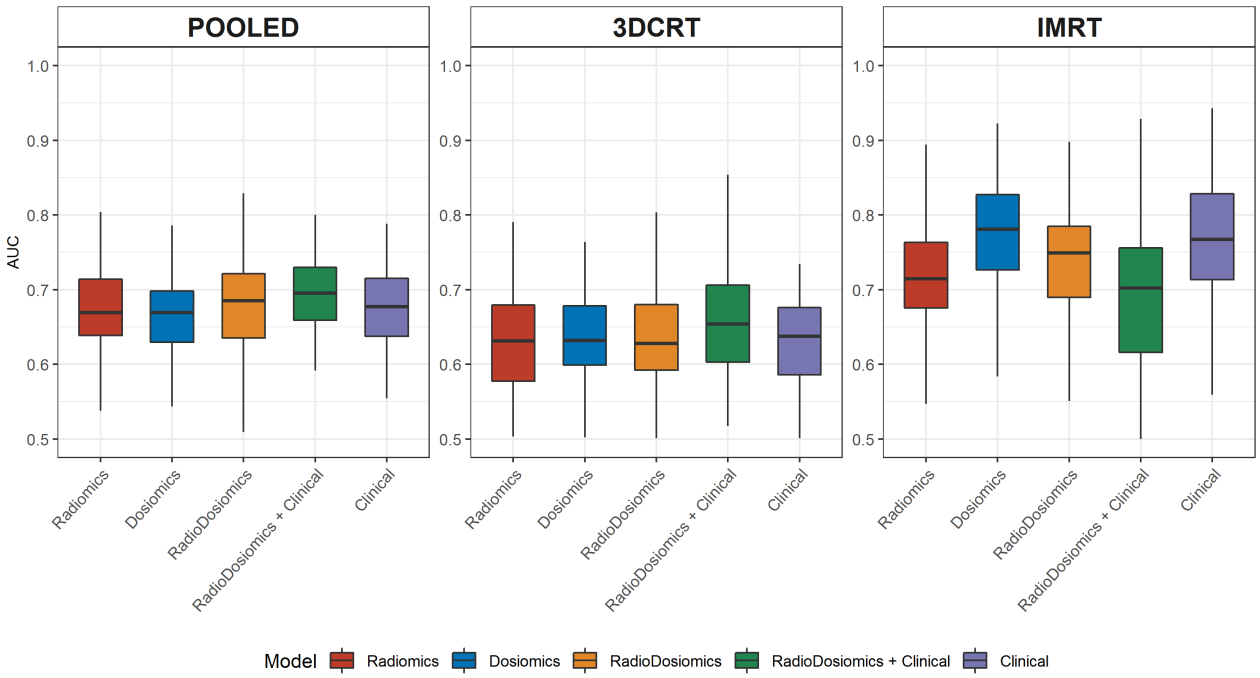

Figure A-15: Logistic classifier performance of the bilateral ROI-V5 radiomics, dosiomics and the combined models plotted along with the clinical model

Logistic Classifier for ROI-V20 Ipsilateral Features

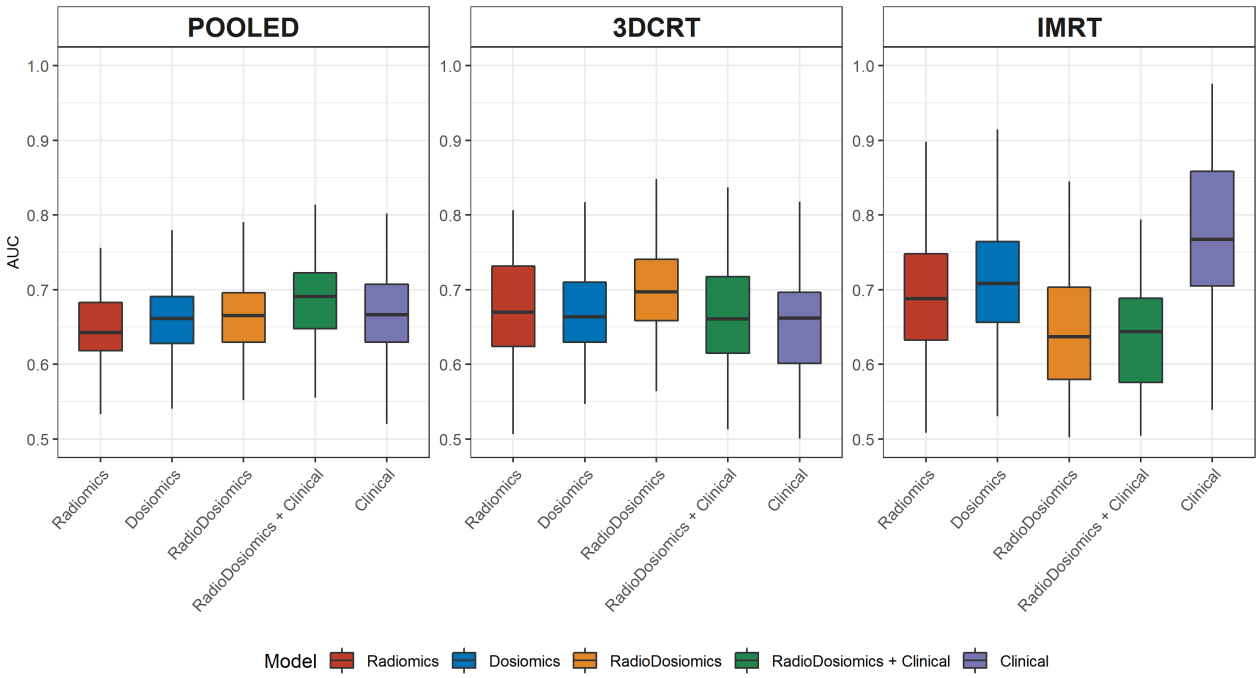

Figure A-16: Logistic classifier performance of the ipsilateral ROI-V20 radiomics, dosiomics and the combined models plotted along with the clinical model

Logistic Classifier for ROI-V5 Ipsilateral Features

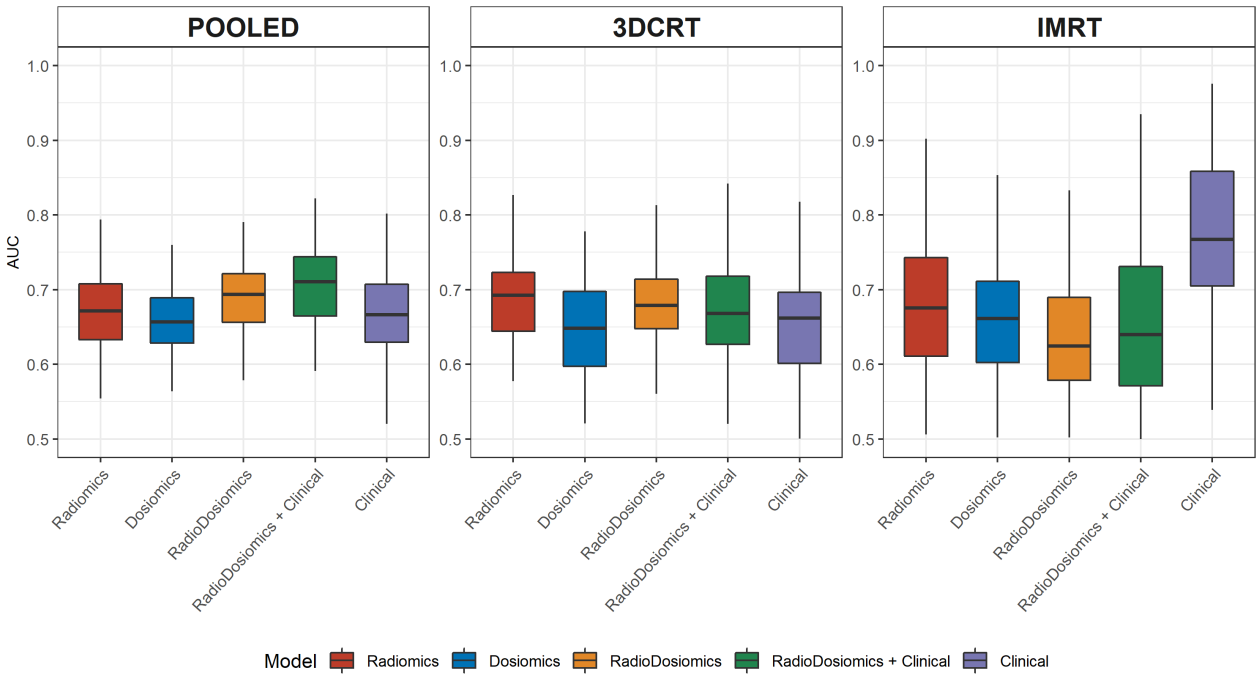

Figure A-17: Logistic classifier performance of the ipsilateral ROI-V5 radiomics, dosiomics and the combined models plotted along with the clinical model

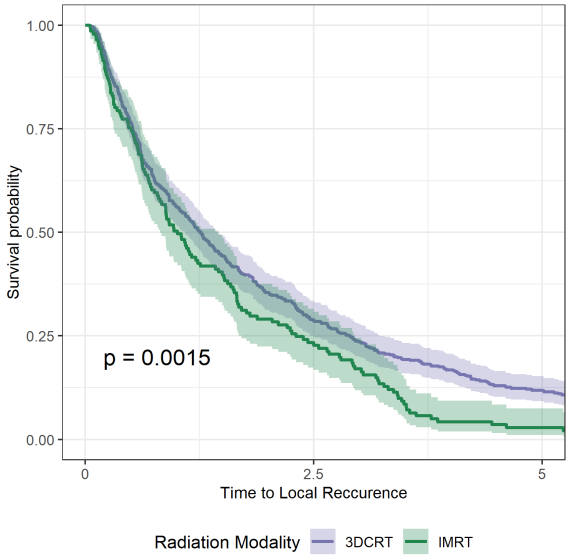

Figure A-18: Locoregional failure rates. A significant number of patients who were treated with 3DCRT and IMRT reccured within 1 year

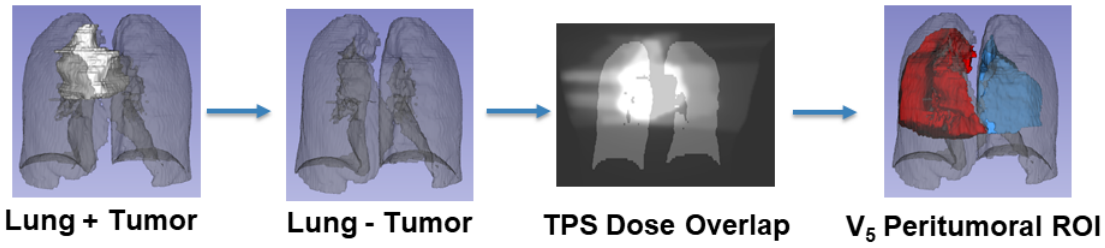

Figure A-19: The process of generating ROI-V5 by using the anatomical (CT) and dose distribution (TPS). The tumor subtracted pulmonary structures were overlayed with the dose distribution to retain a binary pulmnonary volume that received $\geq$ 5 Gy

Appendix B

Tables

Feature Class	Original	LoG (5x)	Wavelet (15x)
Shape	14		
First Order Statistics	18	90	288
Gray Level Co-occurrence Matrix (GLCM)	24	120	360
Gray Level Size Zone Matrix (GLSZM)	16	80	240
Gray Level Run Length Matrix (GLRLM)	16	80	240
Neighbouring Gray Tone Difference Matrix (NGTDM)	5	25	75
Gray Level Dependence Matrix (GLDM)	14	70	210

Table B.1: Breakdown of 1967 Radiomics & Dosiomics features listing the enabled feature classes and corresponding number of features extracted from the different image types i.e Original (unfiltered) image, Laplacian of Gaussian (LoG) filtered image, and Wavelet filtered image

	ROI-V5			ROI-V20		
Features	**Pooled**	**IMRT**	**3DCRT**	**Pooled**	**IMRT**	**3DCRT**
Radiomics	1483			1336		
Dosiomics	1222	1072	1266	1185	1119	1172

Table B.2: Number of robust radiomics and dosiomics features (ICC$\geq$0.80). Radiomics robustness was evaluated on the RIDER Test-Retest dataset (n=22) and the dosiomics robustness was evaluated on the dose images of the excluded cases (n=102).

Datasets		
Feature Type	**Lung Side**	**ROI**
Radiomics	Ipsilateral	ROI-V20
Dosiomics	Ipsilateral	ROI-V20
Radiomics	Ipsilateral	ROI-V5
Dosiomics	Ipsilateral	ROI-V5
Radiomics	Bilateral	ROI-V20
Dosiomics	Bilateral	ROI-V20
Radiomics	Bilateral	ROI-V5
Dosiomics	Bilateral	ROI-V5

Table B.3: Feature Datasets

	Total (n = 596)	3DCRT (n= 452)	IMRT (n = 144)	p Value
Age	$65.0\,yrs\,,(28-93)$	$65.0\,yrs\,,(28-93)$	$64.6\,yrs\,,(32-90)$	0.72
Gender				0.39
Female	299 (50.2%)	222 (49.1%)	77 (53.5%)	
Male	297 (49.8%)	230 (50.9%)	67 (46.5%)	
Smoking				0.14
Never	49 (8.2%)	38 (8.4%)	11 (7.6%)	
Current	239 (40.1%)	171 (37.8%)	68 (47.2%)	
Former	308 (52.7%)	243 (53.8%)	65 (45.1%)	
Radiation Pneumonitis				0.02
No	523 (87.8%)	405 (89.6%)	118 (81.9%)	
Yes	73 (12.2%)	47 (10.4%)	26 (18.1%)	
Performance Status				0.10
0	204 (34.2%)	144 (31.9%)	60 (41.7%)	
$1-2$	379 (63.6%)	298 (65.9%)	81 (56.2%)	
$3-4$	13 (2.2%)	10 (2.2%)	3 (2.1%)	
Treatment Modality				0.01
RTOnly	36 (6%)	29 (6.4%)	7 (4.9%)	
ChemoRT	354 (59.4%)	252 (55.8%)	102 (70.8%)	
ChemoRT + Surg	145 (24.3%)	119 (26.3%)	26 (18%)	
Surg + ChemoRT	61 (10.2%)	52 (11.5%)	9 (6.3%)	
Overall Stage				0.01
IIA	34 (5.7%)	30 (6.6%)	4 (2.8%)	
IIB	35 (5.9%)	26 (5.8%)	9 (6.2%)	
IIIA	332 (55.7%)	263 (58.2%)	69 (47.9%)	
IIIB	195 (32.7%)	133 (29.4%)	62 (43.1%)	
Histology				0.44
Adenocarcinoma	266 (44.6%)	206 (45.5%)	60 (41.7%)	
Squamous Cell Carcinoma	191 (32.0%)	144 (31.8%)	47 (32.6%)	
Large Cell Carcinoma	99 (16.6%)	74 (16.4%)	25 (17.4%)	

Table B.4: Patient Characterstics, Treatment Information, and Outcomes

	3DCRT (n= 452)	IMRT (n = 144)	p Value
Mean PTV Volume (cc)	374.1	517.9	1.5×10^{-10}
RT Dose Delivered			0.08
$\leq 54Gy$	126 (27.9%)	25 (17.4%)	
$\leq 60Gy$	80 (17.7%)	28 (19.4%)	
$\leq 66Gy$	198 (43.8%)	73 (50.7%)	
$> 66Gy$	48 (10.6%)	18 (12.5%)	
DVH Metrics			
$Bilateral\, V_{20}$	24.0	26.3	3.4×10^{-5}
$Bilateral\, V_5$	40.6	49.8	8.4×10^{-14}
$MLD\,[Gy]$	13.5	15.2	1.1×10^{-6}
Mean ROI Volume [cc]			
$Bilateral\, V_{20}$	778.6	835.7	0.08
$Bilateral\, V_5$	1343.4	1612.8	2.6×10^{-6}
$Ipsilateral\, V_{20}$	646.9	706.7	0.04
$Ipsilateral\, V_5$	909.9	1048.6	6.1×10^{-4}

Table B.5: Dosimetric Characterstics

Clinical Predictor	Variable Encoding
Age	continuous
Gender	binary
Smoking	binary
Pack Years	continuous
Concurrent Chemo	binary
Performance status (ps)	categorical
Mean Lung Dose (MLD)	continuous
V20	continuous
V5	continuous
Surgery	binary
Treatment Modality	categorical

Table B.6: Encoding type for all the clinical variables considered in the *Clinical* model

	Median AUC			Range		
Bilateral Model	Pooled	IMRT	3DCRT	Pooled	IMRT	3DCRT
Clinical	0.72	0.73	0.69	0.54-0.84	0.55-0.85	0.53-0.82
Radiomics	0.60	0.67	0.66	0.51-0.72	0.50-0.89	0.51-0.81
Dosiomics	0.66	0.70	0.66	0.51-0.80	0.51-0.89	0.51-0.79
Radiomics + Dosiomics	0.71	0.72	0.76	0.58-0.81	0.54-0.90	0.60-0.89
Radiomics + Dosiomics + Clinical	0.76	0.79	0.77	0.67-0.86	0.63-0.93	0.59-0.90

Table B.7: Test AUC for bilateral ROI-V20 random forest models iterated 100 times into 70% train, 30% test with 10-fold cross validation and SMOTE resampling to address class imbalance.

	Median AUC			Range		
Bilateral Model	Pooled	IMRT	3DCRT	Pooled	IMRT	3DCRT
Clinical	0.72	0.73	0.69	0.54-0.84	0.55-0.85	0.53-0.82
Radiomics	0.60	0.72	0.58	0.50-0.70	0.50-0.95	0.50-0.72
Dosiomics	0.59	0.71	0.59	0.50-0.80	0.50-0.90	0.50-0.74
Radiomics + Dosiomics	0.65	0.77	0.63	0.55-0.77	0.56-0.94	0.50-0.77
Radiomics + Dosiomics + Clinical	0.73	0.81	0.72	0.59-0.81	0.62-0.95	0.56-0.86

Table B.8: Test AUC for bilateral ROI-V5 random forest models iterated 100 times into 70% train, 30% test with 10-fold cross validation and SMOTE resampling to address class imbalance.

	Median AUC			Range		
Ipsilateral Model	Pooled	IMRT	3DCRT	Pooled	IMRT	3DCRT
Clinical	0.73	0.74	0.69	0.61-0.82	0.56-0.93	0.54-0.85
Radiomics	0.60	0.68	0.64	0.50-0.69	0.51-0.86	0.51-0.79
Dosiomics	0.65	0.68	0.68	0.52-0.73	0.52-0.86	0.53-0.85
Radiomics + Dosiomics	0.68	0.71	0.75	0.53-0.81	0.50-0.93	0.61-0.87
Radiomics + Dosiomics + Clinical	0.75	0.77	0.77	0.61-0.83	0.58-0.93	0.61-0.89

Table B.9: Test AUC for ipsilateral ROI-V20 random forest models iterated 100 times into 70% train, 30% test with 10-fold cross validation and SMOTE resampling to address class imbalance.

	Median AUC			Range		
Ipsilateral Model	Pooled	IMRT	3DCRT	Pooled	IMRT	3DCRT
Clinical	0.73	0.74	0.69	0.61-0.82	0.56-0.93	0.54-0.85
Radiomics	0.64	0.69	0.67	0.52-0.78	0.50-0.86	0.51-0.81
Dosiomics	0.59	0.65	0.65	0.50-0.69	0.50-0.88	0.50-0.77
Radiomics + Dosiomics	0.67	0.72	0.77	0.52-0.80	0.53-0.91	0.61-0.92
Radiomics + Dosiomics + Clinical	0.75	0.79	0.77	0.66-0.83	0.61-0.92	0.61-0.91

Table B.10: Test AUC for ipsilateral ROI-V5 random forest models iterated 100 times into 70% train, 30% test with 10-fold cross validation and SMOTE resampling to address class imbalance.

	Median AUC			Range		
Bilateral Model	Pooled	IMRT	3DCRT	Pooled	IMRT	3DCRT
Clinical	0.68	0.77	0.64	0.55-0.79	0.56-0.94	0.50-0.83
Radiomics	0.64	0.69	0.67	0.52-0.77	0.50-0.89	0.50-0.81
Dosiomics	0.65	0.73	0.60	0.53-0.79	0.53-0.91	0.50-0.76
Radiomics + Dosiomics	0.66	0.69	0.70	0.50-0.81	0.50-0.90	0.52-0.86
Radiomics + Dosiomics + Clinical	0.67	0.66	0.66	0.52-0.78	0.50-0.9	0.52-0.85

Table B.11: Test AUC for bilateral ROI-V20 logistic classification models iterated 100 times into 70% train, 30% test with 10-fold cross validation and SMOTE resampling to address class imbalance.

	Median AUC			Range		
Bilateral Model	Pooled	IMRT	3DCRT	Pooled	IMRT	3DCRT
Clinical	0.68	0.77	0.64	0.55-0.79	0.56-0.94	0.50-0.83
Radiomics	0.67	0.71	0.63	0.54-0.84	0.50-0.90	0.50-0.79
Dosiomics	0.67	0.78	0.63	0.52-0.79	0.58-0.91	0.50-0.76
Radiomics + Dosiomics	0.68	0.75	0.63	0.51-0.83	0.51-0.94	0.50-0.80
Radiomics + Dosiomics + Clinical	0.70	0.70	0.65	0.53-0.86	0.50-0.93	0.52-0.85

Table B.12: Test AUC for bilateral ROI-V5 logistic classification models iterated 100 times into 70% train, 30% test with 10-fold cross validation and SMOTE resampling to address class imbalance.

	Median AUC			Range		
Ipsilateral Model	Pooled	IMRT	3DCRT	Pooled	IMRT	3DCRT
Clinical	0.67	0.77	0.66	0.52-0.80	0.54-0.98	0.50-0.82
Radiomics	0.64	0.69	0.67	0.51-0.76	0.51-0.90	0.51-0.81
Dosiomics	0.66	0.71	0.66	0.54-0.78	0.53-0.93	0.55-0.82
Radiomics + Dosiomics	0.67	0.64	0.70	0.55-0.79	0.50-0.84	0.56-0.85
Radiomics + Dosiomics + Clinical	0.69	0.64	0.66	0.54-0.81	0.50-0.79	0.51-0.84

Table B.13: Test AUC for ipsilateral ROI-V20 logistic classification models iterated 100 times into 70% train, 30% test with 10-fold cross validation and SMOTE resampling to address class imbalance.

	Median AUC			Range		
Ipsilateral Model	Pooled	IMRT	3DCRT	Pooled	IMRT	3DCRT
Clinical	0.67	0.77	0.66	0.52-0.80	0.54-0.98	0.50-0.82
Radiomics	0.67	0.68	0.69	0.55-0.79	0.51-0.90	0.58-83
Dosiomics	0.66	0.66	0.65	0.51-0.78	0.50-0.88	0.52-0.78
Radiomics + Dosiomics	0.69	0.62	0.68	0.53-0.79	0.50-0.83	0.56-0.84
Radiomics + Dosiomics + Clinical	0.71	0.64	0.67	0.52-0.82	0.50-0.93	0.52-0.84

Table B.14: Test AUC for ipsilateral ROI-V5 logistic classification models iterated 100 times into 70% train, 30% test with 10-fold cross validation and SMOTE resampling to address class imbalance.

Bibliography

[1] Hugo J.W.L. Aerts, Emmanuel Rios Velazquez, Ralph T.H. Leijenaar, Chintan Parmar, Patrick Grossmann, Sara Cavalho, Johan Bussink, René Monshouwer, Benjamin Haibe-Kains, Derek Rietveld, Frank Hoebers, Michelle M. Rietbergen, C. René Leemans, Andre Dekker, John Quackenbush, Robert J. Gillies, and Philippe Lambin. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nature Communications*, 5, 2014.

[2] Houda Bahig, Edith Filion, Toni Vu, Jean Chalaoui, Louise Lambert, David Roberge, Michel Gagnon, Bernard Fortin, Dominic Béliveau-Nadeau, Dominique Mathieu, and Marie Pierre Campeau. Severe radiation pneumonitis after lung stereotactic ablative radiation therapy in patients with interstitial lung disease. *Practical Radiation Oncology*, 6(5), 2016.

[3] Y Benjamini and Y Hochberg. Benjamini-1995.pdf. *Journal of the Royal Statistical Society B*, 57(1):289–300, 1995.

[4] John Boyle, Brad Ackerson, Lin Gu, and Chris R. Kelsey. Dosimetric advantages of intensity modulated radiation therapy in locally advanced lung cancer. *Advances in Radiation Oncology*, 2(1):6–11, 2017.

[5] Julie Brahmer, Karen L. Reckamp, Paul Baas, Lucio Crinò, Wilfried E.E. Eberhardt, Elena Poddubskaya, Scott Antonia, Adam Pluzanski, Everett E. Vokes, Esther Holgado, David Waterhouse, Neal Ready, Justin Gainor, Osvaldo Arén Frontera, Libor Havel, Martin Steins, Marina C. Garassino, Joachim G. Aerts, Manuel Domine, Luis Paz-Ares, Martin Reck, Christine Baudelet, Christopher T. Harbison, Brian Lestini, and David R. Spigel. Nivolumab versus docetaxel in advanced squamous-cell non-small-cell lung cancer. *New England Journal of Medicine*, 373(2), 2015.

[6] Freddie Bray, Jacques Ferlay, Isabelle Soerjomataram, Rebecca L. Siegel, Lindsey A. Torre, and Ahmedin Jemal. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 68(6), 2018.

[7] S. Carvalho, R.T.H. Leijenaar, E.G.C. Troost, W. van Elmpt, J.-P. Muratet, F. Denis, D. De Ruysscher, H.J.W.L. Aerts, and P. Lambin. Early variation of FDG-PET radiomics features in NSCLC is related to overall survival - the "delta radiomics" concept. *Radiotherapy and Oncology*, 118, 2016.

[8] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16(Sept. 28):321–357, 2002.

[9] Sugama Chicklore, Vicky Goh, Musib Siddique, Arunabha Roy, Paul K. Marsden, and Gary J.R. Cook. Quantifying tumour heterogeneity in 18F-FDG PET/CT imaging by texture analysis, 2013.

[10] Stephen G. Chun, Chen Hu, Hak Choy, Ritsuko U. Komaki, Robert D. Timmerman, Steven E. Schild, Jeffrey A. Bogart, Michael C. Dobelbower, Walter Bosch, James M. Galvin, Vivek S. Kavadi, Samir Narayan, Puneeth Iyengar, Clifford G. Robinson, Raymond B. Wynn, Adam Raben, Mark E. Augspurger, Robert M. MacRae, Rebecca Paulus, and Jeffrey D. Bradley. Impact of intensity-modulated radiation therapy technique for locally advanced non-small-cell lung cancer: A secondary analysis of the NRG oncology RTOG 0617 randomized clinical trial. *Journal of Clinical Oncology*, 35(1):56–62, 2017.

[11] Thibaud P. Coroller, Vishesh Agrawal, Elizabeth Huynh, Vivek Narayan, Stephanie W. Lee, Raymond H. Mak, and Hugo J.W.L. Aerts. Radiomic-Based Pathological Response Prediction from Primary Tumors and Lymph Nodes in NSCLC. *Journal of Thoracic Oncology*, 12(3):467–476, 2017.

[12] Thibaud P. Coroller, Patrick Grossmann, Ying Hou, Emmanuel Rios Velazquez, Ralph T.H. Leijenaar, Gretchen Hermann, Philippe Lambin, Benjamin Haibe-Kains, Raymond H. Mak, and Hugo J.W.L. Aerts. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. *Radiotherapy and Oncology*, 114(3):345–350, 2015.

[13] Perrine Créquit, Marie Wislez, Jocelyne Fleury Feith, Nathalie Rozensztajn, Laurence Jabot, Sylvie Friard, Armelle Lavole, Valérie Gounant, Julie Fillon, Martine Antoine, and Jacques Cadranel. Crizotinib associated with ground-glass opacity predominant pattern interstitial lung disease a retrospective observational cohort study with a systematic literature review. *Journal of Thoracic Oncology*, 10(8), 2015.

[14] Alexandra Cunliffe, Samuel G. Armato, Richard Castillo, Ngoc Pham, Thomas Guerrero, and Hania A. Al-Hallaq. Lung texture in serial thoracic computed tomography scans: Correlation of radiomics-based features with radiation therapy dose and radiation pneumonitis development. *International Journal of Radiation Oncology Biology Physics*, 91(5):1048–1056, 2015.

[15] Alexandra R. Cunliffe, Samuel G. Armato, Christopher Straus, Renuka Malik, and Hania A. Al-Hallaq. Lung texture in serial thoracic CT scans: Correlation with radiologist-defined severity of acute changes following radiation therapy. *Physics in Medicine and Biology*, 59(18):5387–5398, 2014.

[16] Hansen KC Das C, Lucia MS and Tyler JK. 02-Applications and limitations of radiomics Stephen. *Physiology & behavior*, 176(3):139–148, 2017.

[17] J. S. De Jong, P. J. Van Diest, and J. P.A. Baak. Heterogeneity and reproducibility of microvessel counts in breast cancer. *Laboratory Investigation*, 73(6), 1995.

[18] Myriam Delaunay, Jacques Cadranel, Amélie Lusque, Nicolas Meyer, Valérie Gounaut, Denis Moro-Sibilot, Jean Marie Michot, Judith Raimbourg, Nicolas Girard, Florian Guisier, David Planchard, Anne Cécile Metivier, Pascale Tomasini, Eric Dansin, Maurice Pérol, Marion Campana, Oliver Gautschi, Martin Früh, Jean David Fumet, Clarisse

Audigier-Valette, Sébastien Couraud, Stéphane Dalle, Marie Thérèse Leccia, Marion Jaffro, Samia Collot, Grégoire Prévot, Julie Milia, and Julien Mazieres. Immune-checkpoint inhibitors associated with interstitial lung disease in cancer patients. *European Respiratory Journal*, 50(2), 2017.

[19] R. Fisher, L. Pusztai, and C. Swanton. Cancer heterogeneity: Implications for targeted therapeutics, 2013.

[20] Hubert S. Gabryś, Florian Buettner, Florian Sterzing, Henrik Hauswald, and Mark Bangert. Design and selection of machine learning methods using radiomics and dosiomics for normal tissue complication probability modeling of xerostomia. *Frontiers in Oncology*, 8(MAR):1–20, 2018.

[21] Matthias Gamer, Jim Lemon, Ian Fellows, and Puspendra Singh. Various Coefficients of Interrater Reliability and Agreement. *Http://Cran.R-Project.Org/Web/Packages/Irr/Irr.Pdf*, 2012.

[22] Ghazaleh Ghobadi, Sonja Van Der Veen, Beatrijs Bartelds, Rudolf A. De Boer, Michael G. Dickinson, Johan R. De Jong, Hette Faber, Maarten Niemantsverdriet, Sytze Brandenburg, Rolf M.F. Berger, Johannes A. Langendijk, Robert P. Coppes, and Peter Van Luijk. Physiological interaction of heart and lung in thoracic irradiation. *International Journal of Radiation Oncology Biology Physics*, 84(5), 2012.

[23] Mary V. Graham, James A. Purdy, Bahman Emami, William Harms, Walter Bosch, Mary Ann Lockett, and Carlos A. Perez. Clinical dose-volume histogram analysis for pneumonitis after 3D treatment for non-small cell lung cancer (NSCLC). *International Journal of Radiation Oncology Biology Physics*, 45(2), 1999.

[24] C. Grande, M. J. Villanueva, G. Huidobro, and J. Casal. Docetaxel-induced interstitial pneumonitis following non-small-cell lung cancer treatment, 2007.

[25] Paul R. Graves, Farzan Siddiqui, Mitchell S. Anscher, and Benjamin Movsas. Radiation pulmonary toxicity: From mechanisms to management. *Seminars in Radiation Oncology*, 20(3):201–207, 2010.

[26] Peter G Hawkins, Philip S Boonstra, Stephen Hobson, W Jason, D Hearn, James A Hayman, Randall K Ten Haken, M Martha, Paul Stanton, Gregory P Kalemkerian, Nithya Ramnath, Theodore S Lawrence, Matthew J Schipper, Shruti Jolly, Ann Arbor, East Medical, and Ann Arbor. Radiation-Induced Lung Toxicity in Non-Small-Cell Lung Cancer: Understanding the Interactions of Clinical Factors and Cytokines with the Dose-Toxicity Relationship. 125(1):66–72, 2018.

[27] Andreas Hochstrasser, Gabriel Benz, Markus Joerger, Arnoud Templeton, Martin Brutsche, and Martin Frh. Interstitial pneumonitis after treatment with pemetrexed: A rare event? *Chemotherapy*, 58(1), 2012.

[28] Leora Horn, David R. Spigel, Everett E. Vokes, Esther Holgado, Neal Ready, Martin Steins, Elena Poddubskaya, Hossein Borghaei, Enriqueta Felip, Luis Paz-Ares, Adam Pluzanski, Karen L. Reckamp, Marco A. Burgio, Martin Kohlhäeufl, David Waterhouse, Fabrice Barlesi, Scott Antonia, Oscar Arrieta, Jérôme Fayette, Lucio Crinò, Naiyer Rizvi, Martin Reck, Matthew D. Hellmann, William J. Geese, Ang Li, Anne Blackwood-Chirchir, Diane Healey, Julie Brahmer, and Wilfried E.E. Eberhardt. Nivolumab versus

docetaxel in previously treated patients with advanced non-small-cell lung cancer: Two-year outcomes from two randomized, open-label, phase III Trials (CheckMate 017 and CheckMate 057). *Journal of Clinical Oncology*, 35(35), 2017.

[29] Yanqi Huang, Zaiyi Liu, Lan He, Xin Chen, Dan Pan, Zelan Ma, Cuishan Liang, Jie Tian, and Changhong Liang. Radiomics signature: A potential biomarker for the prediction of disease-free survival in early-stage (I or II) non-small cell lung cancer. *Radiology*, 281(3), 2016.

[30] Varsha Jain and Abigail T. Berman. Radiation pneumonitis: Old problem, new tricks. *Cancers*, 10(7):1–16, 2018.

[31] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to Statistical Learning*. 2000.

[32] Carl J. Johnston, Terry W. Wright, Philip Rubin, and Jacob N. Finkelstein. Alterations in the expression of chemokine mRNA levels in fibrosis- resistant and -sensitive mice after thoracic irradiation. *Experimental Lung Research*, 24(3), 1998.

[33] Jun Konishi, Koichi Yamazaki, Ichiro Kinoshita, Hiroshi Isobe, Shigeaki Ogura, Satoko Sekine, Takashi Ishida, Riou Takashima, Megumi Nakadate, Shyu Nishikawa, Takeshi Hattori, Hajime Asahina, Mikado Imura, Eiki Kikuchi, Junko Kikuchi, Naofumi Shinagawa, Hiroshi Yokouchi, Mitsuru Munakata, Hirotoshi Dosaka-Akita, and Masaharu Nishimura. Analysis of the response and toxicity to gefitinib of non-small cell lung cancer. *Anticancer Research*, 25(1 B), 2005.

[34] Shane P. Krafft, Arvind Rao, Francesco Stingo, Tina Marie Briere, Laurence E. Court, Zhongxing Liao, and Mary K. Martel. The utility of quantitative CT radiomics features for improved prediction of radiation pneumonitis. *Medical Physics*, 45(11):5317–5324, 2018.

[35] Max Kuhn. Package 'caret' - Classification and Regression Training, 2019.

[36] Virendra Kumar, Yuhua Gu, Satrajit Basu, Anders Berglund, Steven A. Eschrich, Matthew B. Schabath, Kenneth Forster, Hugo J.W.L. Aerts, Andre Dekker, David Fenstermacher, Dmitry B. Goldgof, Lawrence O. Hall, Philippe Lambin, Yoganand Balagurunathan, Robert A. Gatenby, and Robert J. Gillies. Radiomics: The process and the challenges. *Magnetic Resonance Imaging*, 30(9), 2012.

[37] Stefan L.S. Kwa, Joos V. Lebesque, Jacqueline C.M. Theuws, Lawrence B. Marks, Mike T. Munley, Gunilla Bentel, Dieter Oetzel, Uwe Spahn, Mary V. Graham, Robert E. Drzymala, James A. Purdy, Allen S. Lichter, Mary K. Martel, and Randall K. Ten Haken. Radiation pneumonitis as a function of mean lung dose: An analysis of pooled data of 540 patients. *International Journal of Radiation Oncology Biology Physics*, 42(1), 1998.

[38] Philippe Lambin, Emmanuel Rios-Velazquez, Ralph Leijenaar, Sara Carvalho, Ruud G.P.M. Van Stiphout, Patrick Granton, Catharina M.L. Zegers, Robert Gillies, Ronald Boellard, André Dekker, and Hugo J.W.L. Aerts. Radiomics: Extracting more information from medical images using advanced feature analysis. *European Journal of Cancer*, 48(4):441–446, 2012.

[39] Bin Liang, Hui Yan, Yuan Tian, Xinyuan Chen, Lingling Yan, Tao Zhang, Zongmei Zhou, Lvhua Wang, and Jianrong Dai. Dosiomics: Extracting 3D spatial features from dose distribution to predict incidence of radiation pneumonitis. *Frontiers in Oncology*, 9(APR):1–7, 2019.

[40] Vincent Liu, Dorothy A. White, Maureen F. Zakowski, William Travis, Mark G. Kris, Michelle S. Ginsberg, Vincent A. Miller, and Christopher G. Azzoli. Pulmonary toxicity associated with erlotinib. *Chest*, 132(3), 2007.

[41] Feiran Lou, Camelia S. Sima, Valerie W. Rusch, David R. Jones, and James Huang. Differences in patterns of recurrence in early-stage versus locally advanced non-small cell lung cancer. *Annals of Thoracic Surgery*, 98(5):1755–1761, 2014.

[42] Raymond H. Mak, Brian M. Alexander, Kofi Asomaning, Rebecca S. Heist, Chen Yu Liu, Li Su, Rihong Zhai, Marek Ancukiewicz, Brian Napolitano, Andrzej Niemierko, Henning Willers, Noah C. Choi, and David C. Christiani. A single-nucleotide polymorphism in the methylene tetrahydrofolate reductase (MTHFR) gene is associated with risk of radiation pneumonitis in lung cancer patients treated with thoracic radiation therapy. *Cancer*, 118(14):3654–3665, 2012.

[43] Carlo C. Maley, Patricia C. Galipeau, Jennifer C. Finley, V. Jon Wongsurawat, Xiaohong Li, Carissa A. Sanchez, Thomas G. Paulson, Patricia L. Blount, Rosa Ana Risques, Peter S. Rabinovitch, and Brian J. Reid. Genetic clonal diversity predicts progression to esophageal adenocarcinoma. *Nature Genetics*, 38(4), 2006.

[44] M. Malvezzi, G. Carioli, P. Bertuccio, P. Boffetta, F. Levi, C. La Vecchia, and E. Negri. European cancer mortality predictions for the year 2019 with focus on breast cancer. *Annals of Oncology*, 30(5):781–787, 2019.

[45] Andriy Marusyk, Vanessa Almendro, and Kornelia Polyak. Intra-tumour heterogeneity: A looking glass for cancer?, 2012.

[46] Yuichi Ozawa, Takefumi Abe, Minako Omae, Takashi Matsui, Masato Kato, Hirotsugu Hasegawa, Yasunori Enomoto, Takeaki Ishihara, Naoki Inui, Kazunari Yamada, Koshi Yokomura, and Takafumi Suda. Impact of preexisting interstitial lung disease on acute, extensive radiation pneumonitis: Retrospective analysis of patients with lung cancer. *PLoS ONE*, 10(10), 2015.

[47] Chintan Parmar, Emmanuel Rios Velazquez, Ralph Leijenaar, Mohammed Jermoumi, Sara Carvalho, Raymond H. Mak, Sushmita Mitra, B. Uma Shankar, Ron Kikinis, Benjamin Haibe-Kains, Philippe Lambin, and Hugo J.W.L. Aerts. Robust radiomics feature quantification using semiautomatic volumetric segmentation. *PLoS ONE*, 9(7):1–8, 2014.

[48] R Core Development Team. R: A language and environment for statistical computing. *Vienna, Austria*, 2019.

[49] Milos Radovic, Mohamed Ghalwash, Nenad Filipovic, and Zoran Obradovic. Minimum redundancy maximum relevance feature selection approach for temporal gene expression data. *BMC Bioinformatics*, 18(1):1–14, 2017.

[50] Achim Rittmeyer, Fabrice Barlesi, Daniel Waterkamp, Keunchil Park, Fortunato Ciardiello, Joachim von Pawel, Shirish M. Gadgeel, Toyoaki Hida, Dariusz M. Kowalski, Manuel Cobo Dols, Diego L. Cortinovis, Joseph Leach, Jonathan Polikoff, Carlos Barrios, Fairooz Kabbinavar, Osvaldo Arén Frontera, Filippo De Marinis, Hande Turna, Jong Seok Lee, Marcus Ballinger, Marcin Kowanetz, Pei He, Daniel S. Chen, Alan Sandler, and David R. Gandara. Atezolizumab versus docetaxel in patients with previously treated non-small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. *The Lancet*, 389(10066), 2017.

[51] Linda Rossi, Rik Bijman, Wilco Schillemans, Shafak Aluwini, Carlo Cavedon, Marnix Witte, Luca Incrocci, and Ben Heijmen. Texture analysis of 3D dose distributions for predictive modelling of toxicity rates in radiotherapy. *Radiotherapy and Oncology*, 129(3):548–553, 2018.

[52] Hussin A. Rothan and Siddappa N. Byrareddy. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak, 2020.

[53] Debasish F. Roychowdhury, Catherine A. Cassidy, Patrick Peterson, and Michael Arning. A report on serious pulmonary toxicity associated with gemcitabine-based therapy. *Investigational New Drugs*, 20(3), 2002.

[54] Yvan Saeys, Iñaki Inza, and Pedro Larrañaga. A review of feature selection techniques in bioinformatics, 2007.

[55] Markus S. Schröder, Aedín C. Culhane, John Quackenbush, and Benjamin Haibe-Kains. survcomp: An R/Bioconductor package for performance assessment and comparison of survival models. *Bioinformatics*, 27(22):3206–3208, 2011.

[56] G. Sharp, Rui Li, John Wolfgang, G. Chen, Marta Peroni, Maria Francesca Spadea, and Nagarajan Kandasamy. Plastimatch: an open source software suite for radiotherapy image processing. *Proceedings of the XVI'th International Conference on the use of Computers in Radiotherapy (ICCR)*, 2010.

[57] Hidetaka Uramoto and Fumihiro Tanaka. Recurrence after surgery in patients with NSCLC, 2014.

[58] Joost J.M. Van Griethuysen, Andriy Fedorov, Chintan Parmar, Ahmed Hosny, Nicole Aucoin, Vivek Narayan, Regina G.H. Beets-Tan, Jean Christophe Fillion-Robin, Steve Pieper, and Hugo J.W.L. Aerts. Computational radiomics system to decode the radiographic phenotype. *Cancer Research*, 77(21):e104–e107, 2017.

[59] Peter van Luijk, Hette Faber, Harm Meertens, Jacobus M. Schippers, Johannes A. Langendijk, Sytze Brandenburg, Harm H. Kampinga, and Robert P. Coppes. The Impact of Heart Irradiation on Dose-Volume Effects in the Rat Lung. *International Journal of Radiation Oncology Biology Physics*, 69(2), 2007.

[60] Ivan R. Vogelius and Soren M. Bentzen. A literature-based meta-analysis of clinical risk factors for development of radiation induced pneumonitis, 2012.

[61] Ivan S. Vogelius, David C. Westerly, George M. Cannon, Thomas R. MacKie, Minesh P. Mehta, Chikao Sugie, and Søren M. Bentzen. Intensity-modulated radiotherapy might

increase pneumonitis risk relative to three-dimensional conformal radiotherapy in patients receiving combined chemotherapy and radiotherapy: A modeling study of dose dumping. *International Journal of Radiation Oncology Biology Physics*, 80(3):893–899, 2011.

[62] Shulian Wang, Zhongxing Liao, Xiong Wei, Helen H. Liu, Susan L. Tucker, Chao su Hu, Rodhe Mohan, James D. Cox, and Ritsuko Komaki. Analysis of clinical and dosimetric factors associated with treatment-related pneumonitis (TRP) in patients with non-small-cell lung cancer (NSCLC) treated with concurrent chemotherapy and three-dimensional conformal radiotherapy (3D-CRT). *International Journal of Radiation Oncology Biology Physics*, 66(5), 2006.

[63] Robin Wijsman, Frank J.W.M. Dankers, Esther G.C. Troost, Aswin L. Hoffmann, Erik H.F.M. van der Heijden, Lioe Fee de Geus-Oei, and Johan Bussink. Inclusion of Incidental Radiation Dose to the Cardiac Atria and Ventricles Does Not Improve the Prediction of Radiation Pneumonitis in Advanced-Stage Non-Small Cell Lung Cancer Patients Treated With Intensity Modulated Radiation Therapy. *International Journal of Radiation Oncology Biology Physics*, 99(2), 2017.

[64] Shota Yamamoto, Ronald L. Korn, Rahmi Oklu, Christopher Migdal, Michael B. Gotway, Glen J. Weiss, A. John Iafrate, Dong Wan Kim, and Michael D. Kuo. ALK molecular phenotype in non-small cell lung cancer: CT radiogenomic characterization. *Radiology*, 272(2), 2014.

[65] Chengxiang Yi, Yayi He, Haoran Xia, Helin Zhang, and Peng Zhang. Review and perspective on adjuvant and neoadjuvant immunotherapies in NSCLC. *OncoTargets and Therapy*, 12:7329–7336, 2019.

[66] Da Ping Yu, Xu Cheng, Zhi Dong Liu, and Shao Fa Xu. Comparative beneficiary effects of immunotherapy against chemotherapy in patients with advanced NSCLC: Meta-analysis and systematic review. *Oncology Letters*, 14(2):1568–1580, 2017.

[67] Franz Zehentmayr, Brane Grambozov, Julia Kaiser, Gerd Fastner, and Felix Sedlmayer. Radiation dose escalation with modified fractionation schedules for locally advanced NSCLC: A systematic review. *Thoracic Cancer*, 11(6):1375–1385, 2020.

[68] Binsheng Zhao, Leonard P. James, Chaya S. Moskowitz, Pingzhen Guo, Michelle S. Ginsberg, Robert A. Lefkowitz, Yilin Qin, Gregory J. Riely, Mark G. Kris, and Lawrence H. Schwartz. Evaluating variability in tumor measurements from same-day repeat CT scans of patients with non-small cell lung cancer. *Radiology*, 252(1):263–272, 2009.