August 2016

Radiomics for Response Assessment after Stereotactic Radiotherapy for Lung Cancer

Sarah A. Mattonen

The University of Western Ontario

Supervisor
Dr. Aaron Ward

The University of Western Ontario

Joint Supervisor
Dr. David Palma

The University of Western Ontario

Graduate Program in Medical Biophysics

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

© Sarah A. Mattonen 2016

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Medical Biophysics Commons

Recommended Citation
https://ir.lib.uwo.ca/etd/3926
Abstract

Stereotactic ablative radiotherapy (SABR) is a guideline-specified treatment option for patients with early stage non-small cell lung cancer. This high-precision treatment delivers a high dose of radiation to the tumour while avoiding surrounding normal tissue. After treatment, patients are followed up regularly with computed tomography (CT) imaging to determine treatment response. However, benign radiographic changes to the lung known as radiation-induced lung injury (RILI) frequently occur. Due to the large doses delivered with SABR, these changes can mimic the appearance of a recurring tumour and confound response assessment. The objective of this work was to evaluate the accuracy of radiomics, for prediction of eventual local recurrence based on CT images acquired within 6 months of treatment. We calculated quantitative image features within manually delineated regions of common post-SABR changes (consolidation and ground-glass opacity (GGO)). Second-order texture features could predict recurrence in individual patients with an area under the receiver operating characteristic curve (AUC) of 0.8. To eliminate the need for time-consuming manual delineations, a semi-automated graph cuts based segmentation algorithm was adapted for the consolidative regions. A peri-consolidative region, intended to subsample regions of GGO surrounding the consolidative region, was automatically derived and demonstrated consistent classification performance, with non-inferiority to manual delineations. Physician ability to detect timely local recurrence was also measured on CT imaging, and compared with that of the radiomics tool. Within 6 months post-SABR, physicians assessed the majority of images as no recurrence whereas our radiomics system produced an AUC of 0.85 on the same images. These results suggest that radiomics can detect early changes associated with local recurrence that are not typically considered by physicians. Patients with recurrence tend to have increased presence of ground-glass opacity surrounding consolidative changes compared to patients with benign injury at the early follow-up time point. These appearances detected by radiomics may be early indicators of the promotion and progression to local recurrence. This has the potential to lead to a clinically useful computer aided decision support tool based on routinely acquired CT imaging, which could lead to earlier life-saving salvage opportunities for patients with recurrence and fewer unnecessary invasive investigations of patients with only benign injury.
Keywords

Lung Cancer, Stereotactic Radiotherapy, Radiation-Induced Lung Injury, Recurrence, Computed Tomography, Radiomics, Machine Learning, Segmentation, Classification, Observer Performance
Co-Authorship Statement

This thesis is presented in an integrated article format, the chapters of which are based on the following publications that are either published or in preparation for submission. As first author on the peer-reviewed manuscripts, Sarah Mattonen was a significant contributor to all aspects of the studies, manuscript preparation and submission. Specifically, Sarah designed the research questions and was responsible for planning, implementing and analyzing all experiments. She was also responsible for drafting, revising and submission of all manuscripts. All work was performed under the supervision of Dr. Aaron Ward and Dr, David Palma who provided ongoing guidance and were responsible for study conception, defining the research questions, designing and analyzing the experiments; interpreting the results and drafting the manuscript. For each manuscript contained in this thesis, all other co-authors contributed to reviewing and editing the manuscript and their specific contributions are described below.

Chapter 2 is an original research article entitled “Distinguishing radiation fibrosis from tumour recurrence after stereotactic ablative radiotherapy (SABR) for lung cancer: A quantitative analysis of CT density changes,” and was published in *Acta Oncologica* in June 2013. This manuscript was co-authored by Cornelis Haasbeek, Suresh Senan, David Palma, and Aaron Ward. Cornelis Haasbeek and Suresh Senan provided and curated all retrospective images for analysis, as well as provided clinical interpretation of results.

Chapter 3 is an original research article entitled “Early prediction of tumor recurrence based on CT texture changes after stereotactic ablative radiotherapy (SABR) for lung cancer,” and was published in *Medical Physics* in March 2014. This manuscript was co-authored by Cornelis Haasbeek, Suresh Senan, David Palma, and Aaron Ward. Cornelis Haasbeek and Suresh Senan provided and curated all retrospective images for analysis, as well as provided clinical interpretation of results.

Chapter 4 is an original research article entitled “Imaging texture analysis for automated prediction of lung cancer recurrence after stereotactic radiotherapy,” and was published in the *Journal of Medical Imaging* in October 2015. This manuscript was co-authored by Shyama Tetar, Alexander Louie, Suresh Senan, David Palma, and Aaron Ward. Shyama Tetar
manually contoured regions of interest on computed tomography imaging for analysis. Shyama Tetar, Alexander Louie, and David Palma provided longest axial diameter measurements for validation of the segmentation software. All authors contributed to reviewing and editing the manuscript. Suresh Senan provided clinical interpretation and curated all retrospective images for analysis.

Chapter 5 is an original research article entitled “Detection of local cancer recurrence after stereotactic ablative radiotherapy (SABR) for lung cancer: physician performance versus radiomic assessment,” and was published in the International Journal of Radiation Oncology * Biology * Physics in April 2016. This manuscript was co-authored by Carol Johnson, Alexander Louie, George Rodrigues, Mark Landis, Ian Chan, Roya Etemad-Rezai, Timothy Yeung, Suresh Senan, David Palma, and Aaron Ward. Carol Johnson programmed and developed the observer study user interface. Alexander Louie, George Rodrigues, David Palma, Mark Landis, Ian Chan, and Roya Etemad-Rezai assessed all images in the observer study. Timothy Yeung contributed to development of the feature extraction pipeline.
Acknowledgments

First and foremost, I want to thank my supervisors Dr. Aaron Ward and Dr. David Palma for their continued support and guidance. Dr. Aaron Ward, it has truly been a pleasure working with you these past five years. Thank you for taking me on as a graduate student without ever having written a single line of code and having confidence in me to succeed as an aspiring scientist. Your passion and dedication to research is truly inspiring. I am truly grateful for the genuine mentorship you have provided throughout my graduate studies and it has undoubtedly made me a better communicator, innovator, leader, and overall scientist. Dr. David Palma, thank you for your continued direction and support. Your excitement and passion for translational cancer research is unprecedented. Thank you for providing me with countless opportunities to grow as a researcher and experience the clinical aspects of my work. Both of you have created a positive and supportive environment in the Baines Imaging Lab, and I am grateful for the opportunity to be a part of it.

To my advisory committee, Dr. Grace Parraga, thank you for your guidance throughout my graduate work. Your passion for research has inspired me to become a better researcher.

To our collaborators at the VUMC, Dr. Suresh Senan, Dr. Niels Haasbeek, Dr. Shyama Tetar, and Dr. John van Sornsen de Koste; thank you for the continued support of this work and for providing data for analysis, for without which this project would not be possible.

To all my colleagues at the London Regional Cancer Program and London Health Science Centre, it has been an extremely rewarding experience working and interacting with so many passionate and dedicated people. To my collaborators, Dr. Alex Louie, Dr. George Rodrigues, Dr. Ian Chan, Dr. Roya Etemad-Rezai, and Dr. Mark Landis: thank you for taking the time to assist with my project. A special thank you to Dr. Kitty Huang; the opportunity to work with you on many aspects of this work has been truly rewarding.

To the current and past members of the Baines Imaging Lab, thank you for your continued friendship, support and assistance. Dr. Tharindu De Silva, Sachi Elkerton, Dr. Eli Gibson, Wenchao Han, Peter Martin, Mehrnoush Salarian, Dr. Maysam Shahedi, Derek Soetemans, Andrew Warner, and Yiwen Xu: it has truly been a pleasure working with and getting to
know each and every one of you. A special thank you to Carol Johnson: your ongoing assistance with my project has been invaluable and I am grateful for all of our conversations, both research and sports related. To Dr. Timothy Yeung: you created a positive and collaborative environment in the lab and I am truly grateful for your mentorship and assistance with this work. Thank you to all my fellow Medical Biophysics graduate students for creating a truly supportive and friendly environment, and especially Justin Tse, Adam Paish, Matt Lowerison, and Yiwen Xu for you continued friendship.

Most importantly, I want to thank Ryan, my family and friends. To Ryan: thank you for your continued support and reassurance. I wouldn’t have made it through this journey without you by my side and for that I am truly grateful. To my parents: thank you for giving me opportunities and continued support to pursue my dreams in academics. Your continued encouragement and support have been essential to my success, and I am very appreciative for all you have done.

Lastly, I would like to acknowledge and express my utmost appreciation to the sources of funding which have made this work possible. Thank you to the Natural Sciences and Engineering Research Council of Canada, Lawson Health Research Institute, Cancer Care Ontario, the Ontario Institute for Cancer Research, and the Canadian Cancer Society.
Table of Contents

Abstract .................................................................................................................................................. ii
Co-Authorship Statement ....................................................................................................................... iv
Acknowledgments ................................................................................................................................... vi
Table of Contents ............................................................................................................................... viii
List of Tables ........................................................................................................................................ xiii
List of Figures ...................................................................................................................................... xiv
List of Abbreviations .......................................................................................................................... xix
List of Appendices ................................................................................................................................ xxi
Chapter 1 .............................................................................................................................................. 1
  1 Introduction .................................................................................................................................... 1
    1.1 Lung Cancer Epidemiology ....................................................................................................... 1
    1.2 Lung Cancer Presentation .......................................................................................................... 2
      1.2.1 Lung Cancer Screening .................................................................................................... 2
    1.3 Lung Cancer Diagnosis .......................................................................................................... 3
    1.4 Lung Cancer Staging ................................................................................................................. 4
    1.5 Lung Cancer Treatment ............................................................................................................ 5
      1.5.1 Early Stage NSCLC ......................................................................................................... 5
    1.6 Lung Cancer Follow-up .......................................................................................................... 9
      1.6.1 Computed Tomography Imaging ...................................................................................... 9
    1.7 Response Assessment Following SABR ............................................................................... 16
      1.7.1 Response Evaluation Criteria in Solid Tumours .......................................................... 17
      1.7.2 High-Risk CT Features (HRFs) ...................................................................................... 20
      1.7.3 Positron Emission Tomography ...................................................................................... 21
      1.7.4 Other Novel Imaging Methods ...................................................................................... 23
1.7.5 Limitations of Current Studies Assessing Response Post-SABR .......... 24

1.8 Radiomics ............................................................................................................. 25

1.8.1 Workflow ........................................................................................................ 26

1.8.2 Challenges and Limitations ........................................................................ 35

1.8.3 State of the Art in Lung Cancer Radiomics .............................................. 36

1.9 Thesis Hypothesis and Objectives ........................................................................ 39

1.10 Thesis Outline ....................................................................................................... 42

1.10.1 Chapter 2: Distinguishing radiation fibrosis from tumour recurrence after SABR for lung cancer: a quantitative analysis of CT density changes .... 42

1.10.2 Chapter 3: Early prediction of tumour recurrence based on CT texture changes after SABR for lung cancer ................................................. 42

1.10.3 Chapter 4: Imaging texture analysis for automated prediction of lung cancer recurrence after stereotactic radiotherapy .............................. 43

1.10.4 Chapter 5: Detection of local cancer recurrence after SABR for lung cancer: physician performance versus radiomic assessment .............. 43

1.11 References ............................................................................................................. 44

Chapter 2 ........................................................................................................................... 62

2 DISTINGUISHING RADIATION FIBROSIS FROM TUMOUR RECURRENCE AFTER SABR FOR LUNG CANCER: A QUANTITATIVE ANALYSIS OF CT DENSITY CHANGES .......................... 62

2.1 Introduction ........................................................................................................... 62

2.2 Methods and Materials .......................................................................................... 64

2.2.1 Patient Selection and Treatment ....................................................................... 64

2.2.2 Contours ........................................................................................................... 65

2.2.3 Data Analysis .................................................................................................... 67

2.3 Results ................................................................................................................... 68

2.3.1 Pre-treatment lesions ......................................................................................... 70

2.3.2 Post-SABR Measures of Size ........................................................................ 70

2.3.3 Post-SABR Appearance Measures ................................................................... 73
Chapter 3

EARLY PREDICTION OF TUMOUR RECURRENCE BASED ON CT TEXTURE CHANGES AFTER SABR FOR LUNG CANCER

3.1 Introduction

3.2 Materials and Methods

3.2.1 Patient Selection and Imaging

3.2.2 Region of Interest Delineation

3.2.3 Measures of Size

3.2.4 Measures of Imaging Appearance

3.2.5 Data Analysis

3.2.6 RECIST Progressive Disease Criteria for Prediction

3.2.7 Sensitivity of GGO Delineation

3.3 Results

3.3.1 Group Differences

3.3.2 Classification Performance

3.3.3 Correlation of Measures

3.3.4 RECIST Progressive Disease Criteria for Prediction

3.3.5 Sensitivity of GGO Delineation

3.4 Discussion

3.5 References

Chapter 4

IMAGING TEXTURE ANALYSIS FOR AUTOMATED PREDICTION OF LUNG CANCER RECURRENCE AFTER STEREOTACTIC RADIOThERAPY

4.1 Introduction
List of Tables

**Table 1-1**: High-risk features for recurrence prediction on computed tomography (CT) imaging. .................................................................................................................................. 21

**Table 1-2**: A summary of previous studies using radiomics to assess outcomes and response following lung cancer treatment. ............................................................................................ 38

**Table 2-1**: Baseline patient and treatment variables. ............................................................. 69

**Table 2-2**: Baseline pre-treatment imaging characteristics. ................................................. 70

**Table 3-1**: Baseline patient characteristics and treatment variables. .................................. 87

**Table 3-2**: Descriptions and equations of the second order texture features. ..................... 94

**Table 3-3**: Areas under the receiver operating characteristic curve (AUC) for each feature... 107

**Table 3-4**: Spearman rank correlation coefficients and significance values for features within the ground-glass opacity ....................................................................................................... 109

**Table 4-1**: Area under the receiver operating characteristic curve for the top texture feature (entropy) at both time points post-SABR ............................................................................. 138

**Table 4-2**: Inter-operator variability in the semi-automatic consolidative segmentations; all values are reported as the median [interquartile range]. ....................................................... 139

**Table 5-1**: Baseline patient characteristics. .......................................................................... 163

**Table 5-2**: Physician assessment of response post-SABR based on all follow-up images. . 165

**Table 5-3**: Leave-one-out cross-validation results of the radiomic signature compared to physician assessment at 2–5 months post-SABR ............................................................................................ 167

**Table 5-4**: Top selected features at 2–5 months post-SABR based on leave-one-out feature selection (38 folds). ........................................................................................................ 169
List of Figures

**Figure 1-1**: Radiation induced lung injury following a traditional anterior/posterior parallel opposed pair (treatment plan shown in (a)); (b) The resulting benign injury conforms to the treatment portals and is easily distinguished by a straight line.......................... 10

**Figure 1-2**: Classification of post-SABR radiological changes. Adapted from Dahele et al. [66].......................................................................................................................................... 12

**Figure 1-3**: Planning CT image for SABR treatment and subsequent follow-up imaging after radical treatment for early-stage primary lung cancer. ........................................................... 13

**Figure 1-4**: Radiological changes following SABR for an 85-year-old gentleman with biopsy proven adenocarcinoma. This patient received 54 Gy in 3 fractions with the treatment plan shown in (a). Radiological changes are seen (b) where 0m indicates the pre-treatment lesion measuring 2.0 cm. At 3 months post-SABR, there is presence of an enlarged ground-glass and semi-solid opacity measuring 4.3 cm. At 6 months there is interval reduction in size and a decrease in ground-glass opacity, with ongoing reduction in size at 18 months and 3 years post-treatment. ........................................................................................................................ 15

**Figure 1-5**: (a) Pre-treatment CT image of a 79 year old gentlemen treated with stereotactic radiotherapy and (b-d) post-treatment follow-up CT images at 3, 6, and 9 months respectively. Enlargement of the nodule at 6 and 9 months raised suspicion of local recurrence and the patient was sent for surgical resection. (e) A 7 mm diameter tumour-like fibrous scar was observed on pathology, with sharply defined regions of fibrosis surrounding the scar and mild pleural thickening. Adapted from Takeda et al. [77]................................. 16

**Figure 1-6**: Demonstration of RECIST failure in a patient who received stereotactic radiotherapy for stage I NSCLC. Radiation planning scan (A) shows the prescribed dose (red; 54 Gy in 3 fractions), 50% of prescribed dose (orange) and 25% of prescribed dose (yellow). 3 month scan (B) showed a large area of consolidation meeting RECIST criteria for progressive disease, but the patient was observed. Ongoing observation at 6 months (C) and 40 months (D) showed development of fibrosis with no progression. ................................. 19
Figure 1-7: Radiomics involves image acquisition and region of interest delineation. An example CT image and corresponding region of interest are shown in red. Within the region of interest several image features can be extracted, including first-order statistics, second-order texture and size and shape-based features. These features can be used to predict patient outcomes.

Figure 2-1: Manual contours throughout a course of follow-up for a patient whose cancer recurred. The solid lines represent consolidative changes, and the dashed lines represent ground-glass opacity (GGO).

Figure 2-2: The standard deviation (SD) of CT density was used as a basic texture measure. Three examples from GGO regions of patients in this study, showing the differences in texture with the varying SDs: a) 50.1 HU, b) 85.8 HU, c) 123.4 HU.

Figure 2-3: Cumulative size measures of the consolidative regions throughout follow-up post-SABR; all values are the mean ± 95 % CI. a) RECIST and b) 3D volume. * Indicates statistical significance at p ≤ 0.05.

Figure 2-4: Cumulative appearance measures of the ground glass opacity regions throughout follow-up post-SABR; a) Mean [± 95% CI] CT density, and b) standard deviation [± 95% CI] of CT density. * Indicates statistical significance at p ≤ 0.05.

Figure 2-5: Cumulative appearance measures of the consolidative regions throughout follow-up post-SABR; a) Mean [± 95% CI] CT density, and b) standard deviation [± 95% CI] of CT density. * Indicates statistical significance at p ≤ 0.05.

Figure 3-1: Distribution of post-SABR CT scan time points for the patients in this study. Patients marked as recurrences on this plot experienced eventual cancer recurrence, diagnosed later than the imaging time points indicated on this plot.

Figure 3-2: Manual delineations of post-SABR consolidative and ground-glass opacity findings throughout follow-up for a patient with recurrence (A) and radiation-induced lung injury (B). The zero-month (0m) time point indicates the pre-treatment lesion. The solid lines enclose consolidative regions and the dashed lines enclose ground-glass opacity regions.
Figure 3-3: Flow diagram showing analysis performed on each of the 11 features. This workflow was executed 11 times; once for each of the 11 features used in our study (2 size measures of the consolidation regions and 9 appearance measures of the GGO regions). .... 96

Figure 3-4: Sample texture images and their corresponding second-order energy, entropy, and first-order (standard deviation [SD] of densities) texture values. Samples (a-d) show artificial examples and (e-h) show images taken from lung tissue. The plots show the change of each feature in the specified image relative to those in the leftmost image (a) or (e)....... 100

Figure 3-5: (a) Cross validation errors, (b) false positive rates, and (c) false negative rates for 2-fold cross validation. Glyphs indicate mean values, the error bars indicating the standard deviation over the 100 repetitions................................................................. 102

Figure 3-6: (a) All lesions plotted by their energy and entropy values within the GGO at 2–5 months post treatment. Red circles indicate patients with recurrence and blue crosses indicate those with benign injury. Qualitative example images of the patients indicated by arrows are shown in Figure 3-9. (b,c) The corresponding ROC curves for the linear classifier on each feature alone; AUC = 0.809 for energy and AUC = 0.800 for entropy.............. 105

Figure 3-7: (a) All lesions plotted by their solid (consolidative volume) and longest axial diameter at 2–5 months post treatment. Red circles indicate patients with recurrence and blue crosses indicate those with benign injury. (b,c) The corresponding ROC curves for the linear classifier on each feature alone; AUC = 0.716 for longest axial diameter and AUC = 0.652 for 3D volume. ............................................................................................................ 106

Figure 3-8: Sensitivity of the area under the receiver operating characteristic curve (AUC) for prediction of recurrence at 2-5 months post-SABR based on variations in GGO delineation. The vertical dashed line indicates the AUC values for the manual GGO delineation........................................................................................................................................ 111

Figure 3-9: The 2–5 month scans for four lesions in our study. The solid curves enclose consolidative regions and the dotted curves enclose ground-glass opacity (GGO) regions. Images of patients who had eventual cancer recurrence are shown in (a) and (b), and images of patients who developed RILI are shown in (c) and (d). The energy values in the GGO for these lesions progressively increased from (a) through (d). Their locations in feature space
can be seen in Figure 3-6. (a) and (d) are at relative extremes in feature space on the recurrence and RILI sides, respectively, and are qualitatively distinct. (b) and (c) are examples of recurrence and RILI cases that are closer to the expected decision boundary, and are more difficult to distinguish qualitatively.

**Figure 4-1:** The 2–5 month follow-up image for four patients in our study. The solid lines encompass consolidative changes and the dashed lines encompass regions of ground-glass opacity, as delineated by a senior radiation oncology resident. The two lesions which developed benign radiation induced lung injury are shown in (A) and (B), and qualitatively the ground-glass opacity regions have a smooth appearance. (C) and (D) show two lesions that eventually developed cancer recurrence and in these images a variegated texture is visible in the ground-glass opacity regions.

**Figure 4-2:** An overview of our semi-automatic segmentation approach and the methods used in this chapter. On the original CT image, an operator places a RECIST line segment to measure the longest axial diameter (taken in the normal clinical workflow). This was used as initialization for the *OneCut* algorithm to automatically obtain the consolidation segmentation (red). An expansion of this region defined the peri-consolidative region (blue). On the original CT image manually delineated consolidation and ground-glass opacity (GGO) regions were also obtained. Texture analysis was performed in the GGO and semi-automatic peri-consolidative regions, and classification results were compared using 2-fold cross validation errors, false positive rates (FPR), false negative rates (FNR), and area under the receiver operating characteristic curve (AUC).

**Figure 4-3:** The four in-plane spatial relationships used for calculating the gray level co-occurrence matrix (GLCM). The dark central voxel is the reference voxel, and the neighbouring voxels considered in the analysis are shown in gray.

**Figure 4-4:** Classification performance of the manual GGO and semi-automatic peri-consolidative regions for the top-performing texture feature (entropy) at 2–5 months post-SABR. The columns indicate the mean 2-fold cross-validation errors, mean false positive rates (FPR), mean false negative rates (FNR), and the area under the receiver operating characteristic curve (AUC). The whiskers indicated the standard deviation over 1,000
repetitions and each color represents a different operator. The asterisks indicate a statistically significant difference between operators (p < 0.00005). ...................................................... 137

**Figure 4-5:** The pre-treatment lesion used for treatment planning and a 3-month follow-up scan showing the radiological changes post-SABR for a patient with radiation-induced lung injury (1) and recurrence (2). Semi-automated segmentations of the consolidative regions (red) and peri-consolidative regions (blue) surrounding them. Operators 1, 2 and 3 are shown for each case. ................................................................................................................................. 139

**Figure 4-6:** Non-inferiority analysis for the difference between means (semi-automated – manual) for the classification error metrics (A) and area under the receiver operating characteristic curve (AUC) (B). The markers indicate the mean differences and the whiskers represent the 95% confidence intervals. The vertical dashed lines represent an acceptable clinical difference of 5% for all metrics. .................................................................................. 142

**Figure 5-1:** In-house developed user interface in ClearCanvas Workstation 2.0 (Synaptive Medical, Toronto, Canada) for data collection in observer study.......................... 158

**Figure 5-2:** The 3-month post-SABR follow-up image for two patients who developed local recurrence (A and B) and two patients who developed only benign injury (C and D). The corresponding semi-automated consolidative (red) and peri-consolidative (green) segmentations are also shown................................................................. 161

**Figure 5-3:** Inter-observer variability in follow-up recommendations across all follow-up image for patients with (REC) and without (NOREC) local recurrence. ......................... 166

**Figure 5-4:** Results of folded cross-validation at 2–5 months post-SABR using the five-feature radiomic signature.................................................................................. 170

**Figure 5-5:** Example 2–5 month post-SABR scans for four patients who developed local recurrence (A-D) and four who developed only benign injury (E-H). Correctly classified recurrence (A-C) and injury patients (E-G) in LOOCV by the radiomic signature are shown. A false negative (D) and false positive (H) patient are also shown........................................ 171
### List of Abbreviations

<table>
<thead>
<tr>
<th>1D</th>
<th>One-dimensional</th>
</tr>
</thead>
<tbody>
<tr>
<td>2D</td>
<td>Two-dimensional</td>
</tr>
<tr>
<td>3D</td>
<td>Three-dimensional</td>
</tr>
<tr>
<td>4D</td>
<td>Four-dimensional</td>
</tr>
<tr>
<td>AUC</td>
<td>Area under the receiver operating characteristic curve</td>
</tr>
<tr>
<td>BED</td>
<td>Biologically effective dose</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence interval</td>
</tr>
<tr>
<td>COPD</td>
<td>Chronic obstructive pulmonary disease</td>
</tr>
<tr>
<td>CRT</td>
<td>Conformal radiation therapy</td>
</tr>
<tr>
<td>CT</td>
<td>Computed tomography</td>
</tr>
<tr>
<td>CV</td>
<td>Cross validation</td>
</tr>
<tr>
<td>DCE</td>
<td>Dynamic-contrast-enhanced</td>
</tr>
<tr>
<td>DM</td>
<td>Distant metastases</td>
</tr>
<tr>
<td>DSC</td>
<td>Dice similarity coefficient</td>
</tr>
<tr>
<td>FDG</td>
<td>18-fluorodeoxyglucose</td>
</tr>
<tr>
<td>FFDM</td>
<td>Freedom from distant metastases</td>
</tr>
<tr>
<td>FNR</td>
<td>False negative rate</td>
</tr>
<tr>
<td>FPR</td>
<td>False positive rate</td>
</tr>
<tr>
<td>GGO</td>
<td>Ground-glass opacity</td>
</tr>
<tr>
<td>GLCM</td>
<td>Grey level co-occurrence matrix</td>
</tr>
<tr>
<td>GTV</td>
<td>Gross tumour volume</td>
</tr>
<tr>
<td>HRF</td>
<td>High-risk feature</td>
</tr>
<tr>
<td>HU</td>
<td>Hounsfield unit</td>
</tr>
<tr>
<td>ITK</td>
<td>Insight Segmentation and Registration Toolkit</td>
</tr>
<tr>
<td>LOO</td>
<td>Leave-one-out</td>
</tr>
<tr>
<td>LLL</td>
<td>Left lower lobe</td>
</tr>
<tr>
<td>LR</td>
<td>Local recurrence</td>
</tr>
<tr>
<td>LRC</td>
<td>Loco-regional control</td>
</tr>
<tr>
<td>LUL</td>
<td>Left upper lobe</td>
</tr>
<tr>
<td>MAD</td>
<td>Mean absolute boundary distance</td>
</tr>
<tr>
<td>Acronym</td>
<td>Definition</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>MRI</td>
<td>Magnetic resonance imaging</td>
</tr>
<tr>
<td>NLST</td>
<td>National Lung Screening Trial</td>
</tr>
<tr>
<td>NSCLC</td>
<td>Non-small-cell lung cancer</td>
</tr>
<tr>
<td>OS</td>
<td>Overall survival</td>
</tr>
<tr>
<td>PET</td>
<td>Positron emission tomography</td>
</tr>
<tr>
<td>PTV</td>
<td>Planning target volume</td>
</tr>
<tr>
<td>RECIST</td>
<td>Response Evaluation Criteria in Solid Tumours</td>
</tr>
<tr>
<td>RILI</td>
<td>Radiation-induced lung injury</td>
</tr>
<tr>
<td>RLL</td>
<td>Right lower lobe</td>
</tr>
<tr>
<td>RML</td>
<td>Right middle lobe</td>
</tr>
<tr>
<td>ROC</td>
<td>Receiver operating characteristic curve</td>
</tr>
<tr>
<td>ROI</td>
<td>Region of interest</td>
</tr>
<tr>
<td>RT</td>
<td>Radiation therapy</td>
</tr>
<tr>
<td>RUL</td>
<td>Right upper lobe</td>
</tr>
<tr>
<td>SABR</td>
<td>Stereotactic ablative radiotherapy</td>
</tr>
<tr>
<td>SBRT</td>
<td>Stereotactic body radiation therapy</td>
</tr>
<tr>
<td>SCC</td>
<td>Squamous cell carcinoma</td>
</tr>
<tr>
<td>SCLC</td>
<td>Small cell lung carcinoma</td>
</tr>
<tr>
<td>SD</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>SPLC</td>
<td>Second primary lung cancers</td>
</tr>
<tr>
<td>SUV</td>
<td>Standardized uptake value</td>
</tr>
<tr>
<td>SUVmax</td>
<td>Maximum standardized uptake value</td>
</tr>
<tr>
<td>SVM</td>
<td>Support vector machine</td>
</tr>
<tr>
<td>VD</td>
<td>Volume difference</td>
</tr>
<tr>
<td>VMAT</td>
<td>Volumetric modulated arc therapy</td>
</tr>
<tr>
<td>$^{18}$F</td>
<td>Fluorine-18</td>
</tr>
</tbody>
</table>
List of Appendices

APPENDIX A – Permission for Reproduction of Scientific Articles .................................. 193

APPENDIX B – Supplementary Material to Chapter 2...................................................... 204

APPENDIX C – Supplementary Material to Chapter 4..................................................... 206

APPENDIX D – Curriculum Vitae....................................................................................... 210
Chapter 1

Partial contents of this chapter were previously published in two review articles. The first in the Journal of Thoracic Disease: SA Mattonen, K Huang, AD Ward, S Senan, and DA Palma. New techniques for assessing response after hypofractionated radiotherapy for lung cancer. J Thorac Dis, 2014; 6(4):375-86. Permission to reproduce this article was granted by AME Publishing Company and is provided in Appendix A.1.

The second review article was published in The British Journal of Radiology: SA Mattonen, AD Ward, and DA Palma. Pulmonary imaging after stereotactic radiotherapy - does RECIST still apply? Br J Radio, 2016;20160113. Permission to reproduce this article in a PhD thesis is granted by The British Journal of Radiology and is provided in Appendix A.2.

1 Introduction

1.1 Lung Cancer Epidemiology

Lung cancer remains the leading cause of cancer death in Canada. It is estimated that 26,600 Canadians will be diagnosed and 20,900 Canadians will die from lung cancer in 2015 [1]. On average, 1 in 12 Canadian men and 1 in 15 Canadian women are expected to develop lung cancer in his or her lifetime, and 1 in 13 men and 1 in 17 women are expected to die from it [1]. Five-year survival rates for lung cancer are typically poor, around 15%, as the majority of patients are diagnosed with advanced stage disease [2]. Smoking cigarettes, as well as second-hand or passive smoking, are the predominant risk factors for developing lung cancer [3]. Environmental and occupational factors including air pollution and asbestos have been shown to increase the risk of developing a primary lung cancer [3].

Non-small cell lung cancer (NSCLC) arises in the epithelial cells of the lung from the central bronchi to the alveoli. NSCLC accounts for around 80% of all lung cancers and includes adenocarcinoma, squamous cell carcinoma (SCC), and large cell carcinoma...
The majority of peripheral tumours are adenocarcinomas, whereas squamous cell tumours tend to be located centrally [5]. The majority of patients with NSCLC present with advanced stage III or IV disease [6]. The remaining 20% of lung cancers are small cell lung carcinoma (SCLC), which is typically found in smokers and has a very poor prognosis [4].

1.2 Lung Cancer Presentation

Patients with clinical signs and symptoms of lung cancer typically present with cough, hemoptysis, dyspnea, or chest pain [7]. However, these symptoms can be non-specific and detection of new symptoms can be more difficult in patients with co-existing lung diseases including chronic obstructive pulmonary disease. Physicians should be alert to possible lung cancers when patients present with these symptoms, especially if they are smokers, ex-smokers, or over the age of 50 [8]. When lung cancer is suspected, initial imaging with chest radiography and computed tomography (CT) imaging are typically performed. Once a lung nodule is determined, work-up for a suspected lung cancer will include both diagnostic and staging tests.

1.2.1 Lung Cancer Screening

In patients with a high risk of developing lung cancer, including those between the ages of 55 and 74 who have a history of cigarette smoking of at least 30 pack-years, screening has the potential to find lung cancers early. A recent study by the National Lung Screening Trial (NLST) demonstrated that low-dose CT for screening showed a 20% relative reduction in lung cancer mortality compared with chest radiography [9]. However, approximately 98% of all positive CT screening exams are benign and do not
result in a diagnosis of lung cancer, which may result in unnecessary invasive procedures. The Canadian Task Force on Preventive Health Care now recommends low-dose CT lung cancer screening annually for up to 3 years, for Canadians aged 55 to 74 with at least a 30 pack-year smoking history, who currently smoke or have quit within 15 years. [10]. The introduction of low-dose CT screening will result in an increase in detection of stage I NSCLC.

1.3 Lung Cancer Diagnosis

A physical examination and medical history will assist a physician in determining any signs, symptoms and risk factors associated with a primary lung cancer. Medical imaging, including chest radiography, CT, and positron emission tomography (PET) is also used to diagnose a lung cancer. CT can be used to determine the size, shape, and location of the suspicious lung nodule. On CT imaging, a suspicious lung nodule is typically a rounded or irregular region of increased attenuation. The size of the tumour is measured based on the longest axial diameter and is important for staging. Lung nodules can be classified as solid, sub-solid, or ground-glass [11]. Cavitation of a lesion is a frequent finding in SCC. A contrast-enhanced CT is typically performed to determine if the tumour is invading the mediastinum and to differentiate mediastinal lymph nodes from vascular structures. Suspicious lymph nodes in the mediastinum are typically larger than 10 mm in diameter [12].

Once a lung cancer has been suspected on imaging, a biopsy must be performed to confirm a diagnosis of cancer. Flexible bronchoscopy is a standard procedure to obtain tissue samples for diagnosis of central lesions with high sensitivity [13]. A transthoracic image guided biopsy can be performed for peripheral lesions not accessible by
bronchoscopy [14]. The use of CT-guided transthoracic core needle biopsies has risen over the past decade and is currently recommended for peripheral lung nodules [15]. Pathologic assessment of the tissue will distinguish a NSCLC from a SCLC as well as determine the histological subtype. The histological grade of the tumour is also classified, from well differentiated (grade 1) to undifferentiated (grade 4).

1.4 Lung Cancer Staging

After a diagnosis of NSCLC is made, tumour node metastasis (TNM) classification is critical for determining the appropriate treatment. The TNM classification characterizes a lesion according to the primary tumour (T), nodal status (N), and distant metastasis (M). Staging is determined based on TNM data and in NSCLC there are four primary stages (I, II, III and IV) [16]. Early stage lung cancer is considered stage I disease with small tumours (less than 5 cm) contained within the lung that have not spread to the nearby lymph nodes. Cancers which have spread to bronchial or hilar lymph nodes are considered stage II. Stage III disease, or locally advance disease, has spread to mediastinal lymph nodes. Tumours which have metastasized outside of the lung are considered stage IV disease.

Standard CT imaging has the ability to classify the primary tumour. For mediastinal lymph node staging, CT imaging has a sensitivity and specificity of 83% and 82% respectively [12]. For suspicious lymph nodes adjacent to the mediastinum, a mediastinoscopy is the standard approach for obtaining tissue samples. In this approach a mediastinoscope is inserted through a small incision in the throat to sample a lymph node of interest. Significant advances in staging have developed from the introduction of endoscopic ultrasound and endobroncial ultrasound. These techniques are less invasive
than a standard mediastinoscopy and allow for efficient sampling of regional lymph node stations. 18-fluorodeoxyglucose (FDG) PET imaging also has an increasing role in lung cancer staging and for regional nodal staging the detection of distant metastases [17, 18]. FDG-PET images the metabolic uptake of glucose in tissue and allows for the differentiation of normal and malignant tissue [19].

1.5 Lung Cancer Treatment

The treatment of NSCLC is dependent on the stage and patient’s overall health and lung function [20]. Treatment options can include surgery, radiation therapy (RT), and/or chemotherapy. The role of personalized medicine in lung cancer treatment is also emerging where targeted therapies have the potential to improve outcomes in patients with a known gene mutation [17]. For stage I disease, patients are typically treated with surgery or radiation therapy alone. In stage II disease, surgery may be performed followed by adjuvant chemotherapy and radiation therapy. A combination of surgery, chemotherapy, and/or radiation therapy is typically considered for stage III disease [20]. Patients with stage IV disease are generally considered palliative, but may benefit from a combination of the chemotherapy, radiation therapy, and/or targeted agents [17].

1.5.1 Early Stage NSCLC

1.5.1.1 Surgery

Surgery remains the standard treatment option for patients with early-stage (T1/T2 N0) NSCLC. To determine if a patient is eligible for surgery, pulmonary function or lung function tests are completed to measure lung capacity and function. For patients who are then considered medically operable, a lobectomy is the guideline-recommended
treatment option [21, 22]. Wedge or sublobar resection is recommended for patients with small tumours who are at higher risk with a lobectomy. However, several studies have demonstrated sublobar resection has inferior clinical outcomes compared to standard lobectomy [22, 23]. Minimally invasive surgery through video-assisted thoracic surgery has been shown to have fewer post-operative complications with comparable outcomes to traditional thoracotomy [24].

The most common cause of death following surgical resection of early stage disease is tumour recurrence. Therefore long term survival is moderate, with five-year survival rates in stage I NSCLC following surgery ranging between 55-80% [25-27]. Rates of recurrence following surgical resection range from 30-55%, with most patients failing at distant sites [28]. Local control rates following surgical resection are quite favorable; greater than 90% in many studies [29].

1.5.1.2 Stereotactic Ablative Radiotherapy

Stereotactic ablative radiotherapy (SABR) has become a standard treatment option for patients with early-stage (T1/T2 N0) NSCLC who refuse surgery or are considered medically inoperable [21, 30]. The use of SABR, which is also known as stereotactic body radiation therapy (SBRT), for curative-intent treatment of NSCLC has been rapidly increasing over the last decade [31]. SABR differs from conventional radiotherapy techniques in that it delivers high doses per fraction (approximately 7.5–18 Gy per fraction versus 2 Gy per fraction) over a shorter treatment time (typically 3–8 fractions over 1–2 weeks versus 20–30 fractions over 4–6 weeks). Evidence suggests that a biologically effective dose (BED) in excess of 100 Gy_{10} is required for optimal local control [32]. These high doses are achievable with the use of highly conformal treatment
plans, which include precise planning, targeting, and treatment delivery. SABR is arguably one of the largest medical breakthroughs in the curative treatment of early stage NSCLC in the last two decades, with improved population-based survival rates demonstrated after the implementation of SABR [33-35].

In addition to treatment of primary lung cancer, the use of SABR has also been rapidly increasing for oligometastatic disease [36, 37]. Several single-institution studies have demonstrated high rates of local control, with favorable comparisons to surgery in overall survival outcomes [38, 39]. However, for colorectal cancer, rates of local control after SABR may be lower than other histologies, approximately 70–80% [39]. The impact of SABR on overall survival in patients with oligometastatic disease is currently being evaluated in a randomized trial [40].

1.5.1.2.1 Comparison to Surgery

The effectiveness of SABR for local tumour control has been well established. Reported three-year local control rates often exceed 90% [29, 41]. SABR outcomes appear not only superior to more fractionated regimens [42] (e.g. 55 Gy in 20 fractions [43]), but are comparable to standard surgical resection, as supported by retrospective, single- or multi-institution, and modeling studies, with the largest single-institution retrospective study reporting a 5-year local control rate of 89.5% [29, 44, 45]. Three randomized trials comparing resection vs. SABR have closed due to poor accrual. A pooled analysis of the accrued patients from two trials has been completed, and although the sample size was small, results showed the two treatment options to be comparable [41]. SABR was better tolerated (10% grade 3 toxicity with SABR vs. 44% grade 3–4 toxicity with surgery), with better post-treatment quality of life [46]. SABR achieved
better overall survival than surgery (3-year OS 95% vs. 79%, p=0.037); however, larger studies are needed to confirm these findings.

In high-risk patients with severe pulmonary comorbidities, SABR offers comparable rates of local control without the attendant short-term mortality risks of surgery [47]. In the operable patient population, promising outcomes are reported by two prospective clinical trials: RTOG 0618, reporting a primary tumour failure rate of 7.7% [48], and JCOG 0403, reporting a preliminary 3-year tumour control rate of 86% [49]. For institutions without the capability to deliver SABR, other hypofractionated regimens can also achieve reasonable local control at early time-points: a recent Canadian multicenter study of hypofractionated RT delivering 60 Gy in 15 fractions (BED of 75 Gy_{10}) achieved a two-year local control rate of 88% [50].

1.5.1.2.2 Recurrence Following SABR

Outcomes following SABR are favourable, with recent studies demonstrating 5-year local and regional control rates of 90% and 87% respectively. Local recurrences, typically defined as failure within the treated area, typically manifest at a median time of 15 months post-SABR, but they may present up to 5 years following treatment [51]. Despite high rates of local control, patients still remain at risk of lobar recurrence: the multicenter RTOG 0236 trial demonstrated a 5-year primary tumour recurrence rate of 7%, but an involved lobar recurrence rate of 20% [45, 52]. However, lobar recurrence after SABR may be difficult to distinguish from development of second primary lung cancers (SPLC). Regardless of the classification as recurrence or SPLC, many patients with lobar recurrence can be salvaged with surgical resection [41, 53-55].
Many factors have been identified in the literature as predictive of local control based on Cox multivariable analysis. These include both dose factors, including the biologically effective dose (BED) and minimum planning target volume (PTV) dose, as well as tumour factors including T-stage and gross tumour volume (GTV) size [56-58].

1.6 Lung Cancer Follow-up

After SABR treatment, patients are typically followed with physical examination and CT imaging every 3–6 months for the first 3 years following treatment [51]. Although FDG-PET scans are recommended in lung cancer diagnosis and re-staging, functional imaging currently has a limited role in the evaluation of tumour response [19].

1.6.1 Computed Tomography Imaging

CT imaging plays an important role in lung cancer diagnosis, treatment, and follow-up. Current recommendations for imaging follow-up after SABR are generally based on retrospective evidence and expert opinion, rather than randomized data. Such follow-up serves 3 major goals: detection of local recurrence, detection of regional recurrence that may be amenable to salvage, and detection of new primary lung tumours, which occur at a rate of 2-10% per person-year [51, 59]. Based on the results of the National Lung Screening Trial [60], the American Association for Thoracic Surgery guidelines recommends 4 years of CT follow-up for patients who have undergone treatment for lung cancer and are eligible for additional treatment [61].

1.6.1.1 Radiation-Induced Lung Injury

Following radiotherapy to the lungs, the development of radiographic radiation-induced lung injury (RILI) on CT imaging can occur. Radiation delivered to the tumour
and surrounding lung parenchyma results in radiologic lung injury (pneumonitis and fibrosis) appearing as increased density on CT. The amount of radiation damage to normal tissue varies according to delivered dose, fractionation, dose rate, irradiated volume, and beam arrangement [62]. The appearance and patterns of RILI can also vary across follow-up time intervals. Radiation pneumonitis is typically seen in the acute setting within 6 months of treatment, following which it is classified as fibrosis [63]. From histopathological studies obtained after resection for false-positive imaging studies, these areas of lung injury are made up of a benign mixture of inflammatory cells, fibrocytes and other benign features [64]. Lung injury following traditional 3D-conformal radiation therapy (CRT) is often characterized by straight edges that conform to treatment portals [65], as shown in Figure 1-1.

**Figure 1-1:** Radiation induced lung injury following a traditional anterior/posterior parallel opposed pair (treatment plan shown in (a)); (b) The resulting benign injury conforms to the treatment portals and is easily distinguished by a straight line.
1.6.1.2 Common Appearances after SABR

Following SABR, benign RILI is nearly always present on follow-up CT imaging. Ablative doses of radiation result in radiographic changes appearing as an increased density and opacity on CT in the area of the high-dose region [66, 67]. Such CT changes correlate closely with local delivered dose [68]. The total dose, fractionation, treatment delivery technology, and tumour size are all factors which may affect the degree of radiographic lung injury [68, 69]. Such findings are not unique to lung SABR; they have also been described in other organs treated with stereotactic radiotherapy including brain and liver [70, 71].

The appearance of fibrosis is very common, occurring in 62% of patients within 6 months of treatment (acute) and 91% thereafter (late), as classified by a common classification scheme (Figure 1-2) [66, 67]. This scheme classifies acute radiation pneumonitis into consolidative or ground-glass opacity changes, which can further be subdivided into diffuse (> 5cm) or patchy (≤ 5cm). Late radiation fibrosis can be categorized into modified conventional, mass-like, or scar-like patterns [66, 72]. A modified conventional pattern has been described, defining a fibrosis pattern that is larger than the original tumour size, may be associated with ground-glass opacity, and may include consolidation, volume loss, and bronchiectasis that is similar to or less extensive than conventional radiation fibrosis [66, 73]. These radiographic changes can persist and continue to evolve even after 2 years following treatment.
Morphologic patterns of fibrosis can also vary with treatment type; patients who underwent arc-based SABR had a predicted probability of a modified conventional pattern of 96.3% versus 68.9% for those who underwent fixed-beam treatment [69]. Although such radiologic lung injury occurs in nearly all patients by 2-years [66], only a small minority of patients develop clinical symptoms.

Several studies have examined simple dose-response relationships of HU changes following SABR. Increasing densities on CT post-SABR are seen with larger planning target volumes and longer time post-SABR, and these are most evident in regions receiving doses greater than 20 Gy [68]. Density changes post-SABR have also been shown to linearly increase to doses of 35-40 Gy and then plateau thereafter [68, 74]. The spatial location of fibrosis following SABR is on average 2.6 cm from the GTV position, although displacement of the fibrotic changes of >5 cm can also be observed [75].
1.6.1.2.1 Radiation-Induced Lung Injury versus Recurrence

As a direct result of the highly ablative and conformal doses delivered with SABR, these benign radiographic changes can appear similar to a recurring tumour (Figure 1-3). Against the background of asymptomatic radiation-induced lung injury, accurate assessment of local recurrence is of paramount importance. These changes on CT can result in a major clinical dilemma with respect to accurately distinguishing patients with local recurrence from benign RILI, especially in cases with mass-like changes [76]. Although the classification scheme previously described is used to categorize radiological changes following SABR, it is not used to distinguish recurrence from fibrosis.

**Figure 1-3:** Planning CT image for SABR treatment and subsequent follow-up imaging after radical treatment for early-stage primary lung cancer.

Distinguishing a recurrent tumour from fibrotic lung changes on CT can be challenging for several reasons, as demonstrated in Figure 1-4. Both radiation-induced lung injury and recurrent disease follow a similar temporal course, with lung fibrosis continuing to evolve two years after treatment, during which time, the majority of local recurrences occur [51, 66]. In contrast to lung injury following traditional 3D-CRT, which was often characterized by straight edges that conform to treatment portals (Figure
1-1), the pattern of lung injury on CT following SABR can be mass-like, due to the conformal nature of SABR [66, 76, 77]. Fibrosis may even appear on CT as an enlarging density and therefore can mimic the growth of a local recurrence [77].

Misclassification of a recurrence as “benign fibrosis” can result in a missed window of opportunity for curative-intent salvage treatment. Conversely, misclassification of fibrosis as a recurrence may lead to unnecessary interventions, such as biopsy, imaging, chemotherapy, and even surgery, exposing patients to unnecessary risks and morbidity [64, 77-80]. The ability to accurately assess response is particularly important in light of the changing practice patterns for early stage NSCLC. As a growing number of patients are being treated by SABR [81], this clinical scenario will become more common. The treatment of a fitter patient population may result in a larger proportion of patients who are candidates for salvage treatment in the case of recurrence. Finally, since recent data on potentially operable SABR patients suggest that failure may be higher than in the inoperable SABR cohort (with two-year lobar failure rates in one recent multicenter study [defined as recurrence anywhere in the irradiated lobe] as high as 19.2% [48]), accurate distinction between recurrence and fibrosis to permit early salvage is a pressing clinical problem.
Figure 1-4: Radiological changes following SABR for an 85-year-old gentleman with biopsy proven adenocarcinoma. This patient received 54 Gy in 3 fractions with the treatment plan shown in (a). Radiological changes are seen (b) where 0m indicates the pre-treatment lesion measuring 2.0 cm. At 3 months post-SABR, there is presence of an enlarged ground-glass and semi-solid opacity measuring 4.3 cm. At 6 months there is interval reduction in size and a decrease in ground-glass opacity, with ongoing reduction in size at 18 months and 3 years post-treatment.

Several groups have reported patients with suspicious findings on CT and/or FDG-PET imaging who underwent salvage lung resection to have pathology show no viable tumour cells [77, 80, 82]. Takeda and colleagues demonstrated an enlarging solid
mass on CT follow-up for a patient treated with SABR (as shown in Figure 1-5) [77].
This was suspicious for recurrence and the patient underwent surgical resection, which
demonstrated only a benign fibrotic scar. In most cases, persistent CT findings do not
indicate recurrence: a recent study determining the fate of residual masses after SABR
found that in 50 patients with masses present greater than 1 year following treatment,
only 8 developed local recurrence [83].

Figure 1-5: (a) Pre-treatment CT image of a 79 year old gentlemen treated with
stereotactic radiotherapy and (b-d) post-treatment follow-up CT images at 3, 6, and 9
months respectively. Enlargement of the nodule at 6 and 9 months raised suspicion of
local recurrence and the patient was sent for surgical resection. (e) A 7 mm diameter
tumour-like fibrous scar was observed on pathology, with sharply defined regions of
fibrosis surrounding the scar and mild pleural thickening. Adapted from Takeda et al.
[77]

1.7 Response Assessment Following SABR

With the increase in number of patients receiving SABR for primary lung cancer
or metastatic disease, determining the appropriate follow-up and management of patients
is critical [84]. With a shift toward the use of SABR for patients declining surgery, or
borderline operative candidates, modern cohorts receiving SABR are fit with longer life
expectancies. As a result, surgical or nonsurgical salvage opportunities are available if
failure occurs [53, 54, 85].
1.7.1 Response Evaluation Criteria in Solid Tumours

Response Evaluation Criteria in Solid Tumours (RECIST) is the standard measure of imaging response in oncology. RECIST, first published in 2000, have been widely adopted by many institutions and provide a clear set of guidelines to perform unidimensional measurements for overall evaluation of tumour response. In 2009, the RECIST guidelines were updated to version 1.1 [86] and specific criteria are used to determine tumour response for a target lesion based on measurement of the sum of longest diameters of all target lesions. The baseline sum of longest diameters is used as the reference to characterize response. A complete response denotes the disappearance of all target lesions. A partial response is at least a 30% decrease in sum of longest diameter of the target lesions (reference being the baseline sum of longest diameters). Progressive disease is at least 20% increase in sum of longest diameter of the target lesions (reference being the smallest sum of longest diameter since treatment has started), or the appearance of one or more new lesions. Lastly, stable disease does not have sufficient shrinkage to be considered partial response or sufficient increase in size to be considered progressive disease (less than 20% increase or less than 30% decrease in diameter of the target lesion), again taking as reference the smallest sum of longest diameters since treatment has started.

Response is determined through measurement of the longest diameter of the target lesion within the imaging plane (axial for CT imaging). In the event of isotropic reconstructions, measurements can be made on the reconstructed images in the non-imaging planes. However, since not all radiology sites are capable of producing isotropic reconstructions, caution must be taken to avoid the undesirable situation in which
measurements are taken on different imaging planes at subsequent assessments. It is worth noting that for CT scans of the chest, in which typical slice thicknesses of 5 mm are used; target lesions should have a minimum size of 10 mm to be considered measurable. There are also several other CT image acquisition parameters which should be taken into account for consistency when evaluating lesions using RECIST. These include anatomic coverage, contrast administration, slice thickness, and reconstruction interval which can all impact the evaluation of lesion response [86].

1.7.1.1 Limitations of RECIST

Although RECIST provides a clear set of guidelines for response assessment, they have several limitations [87, 88]. Response assessment based on RECIST relies on physician measurement of lesion diameter. It was been well-described that variability in target lesion diameter exists and this can have an impact on accurately assessing response [89-91]. Inter-observer variability is greater than intra-observer variability, and measurement differences are greatest when there is an irregular edge or spiculated lesion [91]. For consistent measurements, one should consider having a single observer measure the target lesion response across the course of follow-up. The limitation of non-measureable disease in which the lesion diameter is less than 10 mm can be a major limitation after SABR for small lung nodules [87]. The requirement that measurements be taken in the imaging plane can also be a limitation in the context of post-SABR response assessment, since cranio-caudal growth may be a major predictor of recurrence and is measured in the sagittal/coronal plane [92]. Re-evaluation of RECIST 1.1 has been proposed [93].
In the context of response assessment following SABR, the presence of benign fibrotic changes within the high dose region on CT can affect the ability to accurately assess response [35]. When measuring the longest axial diameter of post-SABR changes, it can be unknown if these changes represent viable tumour cells or benign fibrotic tissue. Another limitation of RECIST is in non-spherical lesions which can be difficult to measure. This is specifically important in patients treated with SABR as the appearance and morphology of post-SABR changes can be quite irregular with pleural attachment (as seen in Figure 1-3). This makes accurately determining local lesion response very difficult in the light of significant fibrotic changes following SABR. An example of RECIST failure in a patient treated with SABR is shown in Figure 1-6.

Figure 1-6: Demonstration of RECIST failure in a patient who received stereotactic radiotherapy for stage I NSCLC. Radiation planning scan (A) shows the prescribed dose (red; 54 Gy in 3 fractions), 50% of prescribed dose (orange) and 25% of prescribed dose (yellow). 3 month scan (B) showed a large area of consolidation meeting RECIST criteria for progressive disease, but the patient was observed. Ongoing observation at 6 months (C) and 40 months (D) showed development of fibrosis with no progression.
1.7.2 High-Risk CT Features (HRFs)

A series of HRFs on CT imaging have been identified for detection of local recurrence following SABR. These include the presence of an enlarging opacity, enlargement after one year, sequential enlargement from one scan to the next, bulging margin, linear margin disappearance, and air bronchogram loss [94]. These HRFs were identified based on a systematic review of the literature and then validated in a blinded study of patients with pathologic proof of recurrence [92]. Patients with recurrence were matched 1:2 to patients without local recurrence according to baseline factors. A new HRF of cranio-caudal growth was identified in this cohort. All HRFs were significantly associated with local recurrence and the odds of recurrence increased 4-fold for each additional HRF [92]. A recent validation of these features was performed on an independent patient cohort and demonstrated a bulging margin, linear margin disappearance and craniocaudal growth as the best predictors (Table 1-1) [95]. Combining HRFs was also shown to increase sensitivities and specificities over number of HRFs.

However, not all studies have found all HRFs to be useful. A study by Halpenny et al. examined the predictive value of qualitative CT features for predicting local recurrence following SABR. Eight patients with local recurrence and 83 patients without local recurrence were evaluated for the following signs of local recurrence on CT: a new bulging margin, opacification of air bronchograms, a new or enlarging pleural effusion, a new or enlarging mass, or increased in lung density in the irradiated field. They found that the only feature significantly associated with local recurrence was a new bulging margin at the treatment site [96].
The use of HRFs is subject to limitations. Early detection of local recurrence is difficult, as many require sequential assessments (i.e. sequential enlargement, loss of air bronchograms, loss of linear margin) and may vary depending on frequency of scanning. One HRF cannot be detected until more than a year following treatment. Inter- and intra-observer variability in detecting HRFs is not well-established.

**Table 1-1**: High-risk features for recurrence prediction on computed tomography (CT) imaging.

<table>
<thead>
<tr>
<th>High-risk feature</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Huang et al. [92]</td>
<td>Peulen et al. [95]</td>
</tr>
<tr>
<td>Enlarging opacity at primary site</td>
<td>92</td>
<td>100</td>
</tr>
<tr>
<td>Sequential enlargement</td>
<td>67</td>
<td>62</td>
</tr>
<tr>
<td>Enlargement after 12 months</td>
<td>100</td>
<td>92</td>
</tr>
<tr>
<td>Bulging margin</td>
<td>83</td>
<td>85</td>
</tr>
<tr>
<td>Linear margin disappearance</td>
<td>42</td>
<td>85</td>
</tr>
<tr>
<td>Loss air bronchogram</td>
<td>67</td>
<td>15</td>
</tr>
<tr>
<td>Cranio-caudal growth of ≥ 5 mm and ≥ 20%</td>
<td>92</td>
<td>100</td>
</tr>
</tbody>
</table>

1.7.3 Positron Emission Tomography

The increased opacity on CT is occasionally found with a corresponding increase in metabolic activity on functional imaging in the months following SABR [66, 67]. The use of FDG-PET in the context of response assessment post-SABR has been well studied; however, the data are quite heterogeneous [97-100]. Some studies have shown that the
maximum standardized uptake value (SUVmax) [101-104] and residual SUV uptake 12 weeks post-treatment [105] are strong predictors of local recurrence. Additional work has found that a pre-treatment SUVmax ≥ 5, post-treatment SUVmax ≥ 2 or a reduction in SUVmax < 2.55 were associated with a higher risk of distant failure [106]. Although optimal SUV cutoffs vary across studies, a SUVmax > 5, or greater than the pre-treatment value, appears to be most indicative of recurrence [94, 98, 103-105]. However, many of these studies are subject to type I errors, as multiple SUV cutoffs were assessed for statistical significance.

A major limitation of PET is that inflammatory reaction in areas of the lung receiving high doses from SABR can result in elevated uptake on PET imaging, resulting in false positive findings [107, 108]. These metabolically active FDG-avid lesions may rise transiently immediately post-SABR and persist after 12 months [107-109]. False-positive PET SUVmax readings as high as 7.0 have been reported [64, 110].

Another limitation of the use of FDG-PET imaging is in regard to the standardization of image acquisition across scanners and institutions, which must be considered in the context of these studies [111]. PET is also more costly than standard CT imaging and may not be a routine post-treatment investigation at some institutions. Lack of PET/CT standardization can be an important confounder: measured SUVs can be affected by multiple factors, including technical, physical, and biologic [111]. In order to generalize PET/CT findings, minimum performance or harmonizing standards are needed for many factors including uptake period, patient motion, inflammation, blood glucose level correction, as well as scan acquisition and reconstruction parameters.
A systematic workflow for imaging follow-up post-SABR has been published based on HRFs and SUVmax on PET imaging [92]. This workflow classifies patients as having a low, intermediate, or high risk of recurrence. A more rigorous follow-up might be indicated in patients with a higher likelihood of disease recurrence, including those patients with larger tumours or sub-optimal radiation doses [112, 113]. As more data become available in the management of patients following SABR, applicability of this follow-up recommendation is expected to change.

1.7.4 Other Novel Imaging Methods

Novel imaging modalities may allow for better assessment of treatment responses following SABR. In addition to standard FDG-PET reporting SUV<sub>max</sub> values, functional imaging with additional metrics such as metabolic tumour burden markers may show improvement for assessing response. Preliminary studies have investigated using pre-treatment measures such as metabolic tumour volume and total lesion glycolysis for assessing clinical outcomes after SABR, however further studies with larger samples and follow-up periods are needed [114]. Additional PET tracers such as 18-fluorooazomycin-arabinoside (FAZA) and 18F-fluoromisonidazole (F-MISO) are used for imaging hypoxia in head and neck cancers [115, 116] and could also be investigated for assessing response following hypofractionated radiotherapy.

Perfusion imaging, such as dynamic-contrast-enhanced (DCE) CT or magnetic resonance imaging (MRI) characterizes vascular properties of a tissue and can quantitatively map their spatial distributions. Measures such as blood volume, blood flow, permeability, and mean transit time can be calculated after administration of a contrast agent. Both DCE-CT and DCE-MRI have shown promise as prognostic or
predictive biomarkers in oncology, and their value in assessing response after SABR warrants investigation [114, 117, 118]. CT perfusion imaging has been investigated for response assessment in pulmonary metastases treated with SABR [119]. Although changes in perfusion data were not statistically significant, a qualitative trend consisting of an early increase followed by a decrease in tumour perfusion was noted. Validation on a larger data set is required to determine the role of CT perfusion in response assessment post-SABR. Enhancement patterns have also been investigated following SABR and have shown that patients with recurrence showed a more rapid wash-in and wash-out phenomenon, compared to the continuous enhancement observed in RILI [120].

1.7.5 Limitations of Current Studies Assessing Response Post-SABR

Several potential pitfalls must be considered when evaluating novel imaging modalities for response assessment. First, the gold-standard definition of “recurrence” varies across studies, and many studies use imaging-based definitions of recurrence, rather than pathologic confirmation. Such imaging-based definitions of the endpoint may introduce substantial bias and create a self-fulfilling prophesy: if imaging features are used to define “recurrence” (e.g. sequential growth of lesion) and then the same features are assessed to predict these “recurrences”, their performance may be artificially inflated. The majority of studies include only a small number of biopsy-proven recurrences, and with remainder of patients defined as recurrence an increase in tumour size on successive CT scans [98, 102, 105]. Many also use a modified progression criterion of two consecutive enlargements on CT to define recurrence, therefore suggesting the difficulty in response assessment at an early time point and that the usefulness of PET is limited. Since recurrences are uncommon after SABR, large databases are required to have
sufficient events for analysis, and any new promising markers require robust external validation, since the chances of type I error are high when multiple features are being assessed. Variations in standardization of imaging protocols in both CT and PET studies must assessed for their impact of predictive ability. Finally, post-SABR surgical studies, including registration of digitized histology to CT, would be valuable for correlating imaging findings at the voxel level with true pathologic outcome.

1.8 Radiomics

In contrast to qualitative image assessment, quantitative image feature analysis extracts measurable information from within an image, such as intensities or densities, shape or morphology, or texture. Radiomics is an emerging area of study which aims to extract more information from medical images [121, 122]. The use of radiomics and texture analysis in oncology, and specifically radiation oncology, has been rapidly expanding over the past decade to quantify tumour heterogeneity and predict response [123]. The goal of radiomics is to potentially tailor a patient’s radiotherapy treatment based on predicted response on pre-treatment imaging, or to detect treatment failure at an earlier time point post-treatment [124]. Radiomics look to quantify tumour phenotypes based on a large number of quantitative image features [125]. This can involve the extraction of quantitative image features from regions of interest on either pre- or post-treatment images.

Image feature analysis has emerging roles in general medicine and oncology. Numerous imaging modalities can be used for quantitative image analysis at different body sites, including CT, MRI, ultrasound, and mammography [126, 127]. Applications in oncology include the computer-aided detection or diagnosis of diseases such as breast
and bladder cancer [126, 127]. Texture analysis of the liver has suggested that texture parameters may distinguish high-risk from low-risk colorectal cancer patients [128]. Texture analysis on MRI, CT, and PET has been able to diagnose and characterize tumour heterogeneity for several tumour types and is showing promise in response assessment and as a predictive biomarker [129, 130]. In the thorax, the use of quantitative image feature analysis on CT has been widely investigated in many benign diseases, including characterizing pulmonary infections as well as varying benign lung disease patterns [131-133]. Texture analysis, specifically the product of tumour uniformity and gray-level, has also been correlated with tumour response following chemotherapy in advanced stage NSCLC [134]. The use of CT texture analysis has also been applied to quantify radiation-induced lung damage. Predictive modelling of radiation pneumonitis using texture analysis on CT has been studied following definitive radiation for lung and esophageal cancer [135, 136]. Future work integrating radiomics and genomics (radiogenomics) could aid in characterizing tumour phenotypes and genotypes to associate with outcomes [137]. Many studies have demonstrated the potential of radiomic features to provide additional tumour phenotypic information that may not be visible to the human eye. This information may augment standard clinical or genomic information to provide a more comprehensive analysis of the disease as a whole.

1.8.1 Workflow

An overview of a standard radiomic workflow is shown in Figure 1-1. Following image acquisition, a region of interest (ROI) is defined and within it a series of radiomic image features can be calculated [125]. Following feature extraction, feature selection can be performed to obtain the optimal set of radiomic features. Machine learning using
unsupervised or supervised analysis options, is then performed to summarize information on the data or to build models to predict outcomes or a response variable [138].

**Figure 1-7:** Radiomics involves image acquisition and region of interest delineation. An example CT image and corresponding region of interest are shown in red. Within the region of interest several image features can be extracted, including first-order statistics, second-order texture and size and shape-based features. These features can be used to predict patient outcomes.
1.8.1.1 Segmentation

Radiomic analysis can be performed on any ROI, such as tumour, normal lung, or fibrotic regions; such ROIs can be selected by means of manual, semi-automated, or fully-automated methods. A manual method involves delineation of an ROI by an investigator on each individual slice using imaging software. Manual methods do not require specialized algorithms, but can be tedious and time consuming, and are subject to intra- and inter-observer variability [139]. A semi-automated method requires a smaller amount of user input, and may require a user to initialize the segmentation by selecting a point or region of interest. A fully automated approach requires no user interaction or input and the image is automatically segmented based on a series of predetermined parameters. This makes a fully automated approach quick and reproducible; however the lack of user input or knowledge can be an issue in terms of the reliability. Therefore, semi-automated approaches to segmentation have become increasingly popular as they are reproducible, fast, and require minimal user input or knowledge [140].

Several algorithms exist for semi-automated and fully automated segmentations [141]. In general, segmentation algorithms can be separated into two main categories: low-level and model-based methods. Low-level methods perform the segmentation exclusively based on features in the image. Model-based methods allow for high-level knowledge about the region of interest to be incorporated into the algorithm, including boundary smoothness, appearance, or shape information. The choice of image segmentation algorithm relies on the specific imaging modality being used as well as characteristics of the region to be segmented [142].
Low-level methods include thresholding, clustering methods, and region-growing techniques. The simplest method is thresholding which relies on the selection of a threshold value to convert a grey-scale image into a binary image. Although this is a simple and fast segmentation technique, in some cases selection of the optimal threshold is done manually and this technique does not guarantee object coherency, therefore post-processing may needed. However more advanced thresholding techniques including Otsu’s method have been developed to select the optimal value by minimizing the variance between regions [143]. Clustering methods partition an image into a specified number of clusters according to similar grey-levels. The k-means clustering algorithm is the most common and allows for image segmentation into multiple classes by computing a mean intensity for each class [144]. Region-growing methods work by comparing each pixel with its neighbours and they assume that neighbouring pixels within one region have similar values [145]. In seeded region growing, initialization of the segmentation requires the selection of seeds to mark the object to be segmented. This technique is good for regions with clear edges, however it can be sensitive to image noise, seed point selection and segmentation parameters [146].

Model-based methods can include parametric deformable models and level set algorithms. To achieve the desired segmentation result, parametric deformable models are initialized with a measure of curvature and a contour surrounding the region of interest that is represented parametrically [147]. The algorithm is evolved to minimize an energy/objective function according to image (e.g. edges) and internal (e.g. curvature) terms. These techniques can automatically search for the minimum state and are advantageous as they can incorporate curvature characteristics; however they do require
manual interaction to initialize the model and parameters and can be sensitive to the convergence criteria and local minima [141]. Another limitation of parametric deformable models is their inability to evolve and split into multiple curves to segment multiple disjointed regions of interest. Level set methods address this issue by representing the boundary implicitly. The level set is represented using a signed distance function where zero is the actual contour [148]. Level sets can account for different topologies and provide a direct way to estimate the geometric properties of the structure.

Another class of segmentation methods includes graph partitioning methods [149]. These methods involve having a pixel or group of pixels associated with nodes and edge weights to define similarity between neighbouring pixels. The graph, or image, is then segmented according to different criterion to model specific clusters. Graphs are typically partitioned by finding the minimum cut, by either cutting the minimum number of edges or so that the sum of the cut edges is as small as possible based on a specified cost function [150]. Boykov and Jolly introduced an interactive graph cuts based segmentation for binary image segmentation by minimizing an energy function based on user defined foreground and background samples [151]. Graph cut segmentations can be biased towards producing small contours since it aims to find the minimum cut in the graph. They are also limited to binary segmentations problems: foreground versus background image segmentation. However, it is computationally fast, can be optimized for 3D images, and can achieve a globally optimal solution. Several methods exist for trying to optimize for graph construction, criterion for graph partitioning, and efficient partitioning [152].
1.8.1.2 Feature Extraction

Following region of interest segmentation, qualitative or semantic features of the region of interest can be developed and characterized by expert observers. Such features could include tumour characteristics such as spiculation or pleural attachment. Radiomic image features can also be extracted from the ROI. These features can include first-order statistics based on the distribution of the intensity histogram. These features include things such as the mean, median, standard deviation, variance, skewness, and kurtosis. Intensity refers the brightness of an individual voxel; in CT imaging this can also be described as density and is quantified in Hounsfield Units (HU). HUs measure the attenuation of a material relative to water (HU=0).

CT image texture is a set of more complex measurements which describe local brightness variation or the spatial arrangement of intensities in an image [153, 154]. Second-order texture features take into account the neighbouring relationships of voxels within the region of interest. Extraction of second and third-order texture features can be performed in many ways, including statistical methods, structural methods, model-based methods, and transform-based methods [155]. Statistical texture analysis is the most frequently cited method of texture analysis. This approach describes texture through high-order statistics of an image intensity histogram [155]. These features can include grey-level co-occurrence matrix (GLCM) texture features as well as grey-level run length matrix (GRLM) texture features [156-159]. GLCM features are analyzed by assessing neighbouring voxel pairs; however it can be done with multiple spatial directions and distances. GRLM texture features assess grey-level run lengths in an image which are
defined as consecutive pixels with a specific grey-level. These features are also calculated for a specified spatial direction from a pixel of interest.

Structured approaches obtain descriptions of the spatial relationships of textures through Voronoi tessellation [160]. Model based texture analysis generate an empirical model of each pixel in the image based on neighbouring pixel intensities, and include Markov random fields and fractal models [155]. Transform based texture analysis techniques convert the image into a new form based on the spatial frequency information regarding pixel intensity variations. A common method in radiomics is the use of wavelet features, in which the image is transformed into low and high frequency domains, similar to Fourier analysis, and texture is then calculated on the transformed images [125].

Size and shape based features of the region can also be calculated. Size can be quantified by measures such as longest axial diameter, 3D volume, and surface area [161]. The shape or morphology of a region describes the geometry of the external boundary. Shape-based features can include sphericity, roughness, or spiculation. More advanced quantitative features to characterize shape complexity can also be developed in radiomics to correlate with observed qualitative or semantic image features [162].

1.8.1.3 Machine Learning

Optimal features or sets of features for predictive or prognostic biomarkers must be determined and validated through training and testing on multiple data sets. This can include analyzing individual features alone or a combination of these features together. Machine learning refers to algorithms that can learn and make predictions from data and operate by building models from inputs to make an output based on predictions or
decisions [163]. There are two main types of machine learning: supervised and unsupervised machine learning. In unsupervised machine learning, there are no labels on the training data and the goal is to discover patterns in the data. On the other hand, supervised machine learning is given a set of labels or outputs, and an algorithm tries to determine a rule to map the inputs, or features, to the outputs. Machine learning can also be classified depending on the desired output of the machine learning tool. Classification has a discrete set of labels for the data, for example cancer or non-cancer. In linear regression the outputs are continuous, for example likelihood of cancer. Logistic regression can also be used for predictive analysis and describes the relationship between one dependent binary label and one or more independent variables.

There are several different classification algorithms which can be used in machine learning. Linear classifiers use a linear combination of features to obtain a classification decision [164]. Support vector machine (SVM) classifiers are a maximal margin classifier which can efficiently perform linear classification. One benefit of SVM classifiers is the flexibility to perform non-linear classification. When features are not separable by a linear decision surface, a kernel trick can be performed to achieve non-linear classification. The kernel trick maps the features into a high-dimensional feature space such that is separable by a linear hyperplane. [165]. Decisions trees are another method which can easily visualize and explicitly represent the decision making. Random forests use multiple decision trees to improve classification performance [166]. Artificial neural networks are based on biological neural networks and are used to model complex interactions between inputs and outputs [167].
An emerging area of machine learning is the use of deep learning. Deep learning attempts to model high-level abstractions of data based on multiple processing layers [168]. Many deep learning approaches are based on neural networks with many variations on network architecture to obtain optimal classification performance. A major advantage of deep learning is that it can learn the features from the data itself. However, a major limitation to deep learning is the requirement for extensive amounts of training examples to learn the data.

Machine learning may also include feature selection in which a subset of relevant features are chosen for use in model building. This is done to remove redundant or irrelevant features and to enhance generalizability of the model [169]. There are two main classes of feature selection algorithms: filter and wrapper methods [170]. Filter methods analyze the intrinsic properties of the data by ranking and selecting a subset of the features while ignoring the classifier. These methods are fast and avoid overfitting of the data. Unlike filter methods, wrapper methods evaluate the interaction between a subset of features and can risk overfitting on the data with a high computation time. Some algorithms, including decision trees and random forests, perform feature selection as part of their overall classification operation [166]. Due to the large number of features available as well as the large number of possible combinations of these features, the high-risk of type I error must be recognized when comparisons and cross-validations are performed. As a result, initial exploratory studies in radiomics must be considered hypothesis-generating, and validation on external datasets is crucial.

There are several approaches to evaluate classification performance. In a binary classification problem, the true positive, true negative, false positive and false negative
conditions can all be determined. Based on these values overall accuracy, false positive rate, false negative rate, positive predictive value (precision), sensitivity (recall or true positive rate), and specificity (true negative rate) can be calculated. To determine the trade-off between true and false positive classification, the area under the receiver operating characteristic curve (AUC) can also be calculated. The F1 score can also be measured to convey the balance between precision and recall.

1.8.2 Challenges and Limitations

There are several limitations and challenges when performing radiomic studies [121]. In general, image acquisition should be standardized to minimize any variability between scanners, imaging parameters, or reconstruction techniques. Standardization includes the use of the same scan protocol for imaging acquisition, with consistencies in settings such as kV, mAs, slice collimation, and slice thickness. Breathing instructions and the use of intravenous contrast should also be consistent across all patients, although patients with contra-indications to contrast injection must be noted and studies analyzing the effect of contrast on image feature analysis should be performed. Reconstruction kernels or filters are used to determine image quality of a CT scan and are chosen based on the intended clinical application of the scan. Such decisions are a compromise between spatial resolution and noise, and depending on the organ being scanned, may require a smoother image with less noise or a sharper image with higher noise. Reconstruction kernels should also be consistent across all images and a higher sharpness thorax kernel should be used when available. However, optimal scan parameters and reconstruction kernels must be investigated for the effect of variations among these settings on quantitative image feature analysis.
1.8.3 State of the Art in Lung Cancer Radiomics

The work presented in this thesis was one of the first studies to assess the utility of quantitative image features for response assessment in lung cancer. A summary of other studies using radiomics to assess outcomes and response following lung cancer treatment is shown in Table 1-2. One of the first large scale studies determining the prognostic power of radiomics was by Aerts et al. [125]. They examined 440 intensity, shape, and texture features on CT for predicting outcomes in lung and head and neck cancer patients. They demonstrated the association of radiomic features with clinical data, including TNM descriptors, stage, and histology. They also showed the association of a radiomic signature with overall survival and gene-expression, demonstrating the potential of a radiogenomics analysis to associate quantitative image features with underlying gene-expression patterns. More recent studies have also evaluated complex shape and appearance descriptors to predict overall survival (OS) following lung cancer surgery [162].

Additional studies have determined the potential of radiomics on pre-treatment CT images to predict outcomes in later stage disease (stage II or III) following chemotherapy and radiation therapy [171, 172]. Predictive models incorporating texture features and conventional prognostic factors demonstrated a significant improvement for predicting overall survival (OS), loco-regional control (LRC) and freedom from distant metastases (FFDM) compared to conventional prognostic factors alone [171]. Radiomic features were also found to be associated with distant metastasis (DM), which was validated in an independent validation set [172]. When combining this radiomic signature with a clinical model, prediction of DM was significantly improved, demonstrating the
potential of radiomics to provide additional information regarding tumour phenotype compared to clinical data alone.

For early stage I lung cancers, several studies have investigated the use of pre-treatment CT radiomic features for predicting recurrence after surgery [173, 174]. Emaminejad et al. found that a radiomics and genomic biomarker based classifier could predict recurrence after surgery with an area under the receiver operating characteristic curve (AUC) values of 0.78 and 0.68 respectively [173]. Combining these two classifiers significantly improved performance with an AUC of 0.84. Higher order wavelet features and CT intensities of solid and GGO tumour components demonstrated the ability to predict tumour recurrence with an AUC of 0.8 [174]. These studies demonstrate the potential to assess patients who may be at a higher risk of recurrence pre-treatment to facilitate individualized patient care.
**Table 1-2:** A summary of previous studies using radiomics to assess outcomes and response following lung cancer treatment.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Treatment</th>
<th>Outcome</th>
<th>Modality</th>
<th>Stage</th>
<th>N</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mattonen et al. 2013 [175]</td>
<td>SABR</td>
<td>LR</td>
<td>CT (post-tx)</td>
<td>I</td>
<td>24</td>
<td>Texture</td>
</tr>
<tr>
<td>Mattonen et al. 2014 [176]</td>
<td>SABR</td>
<td>LR</td>
<td>CT (post-tx)</td>
<td>I</td>
<td>24</td>
<td>Radiomics</td>
</tr>
<tr>
<td>Mattonen et al. 2015 [177]</td>
<td>SABR</td>
<td>LR</td>
<td>CT (post-tx)</td>
<td>I</td>
<td>24</td>
<td>Radiomics with semi-automated system</td>
</tr>
<tr>
<td>Emaminejad et al. 2015 [173]</td>
<td>Surgery</td>
<td>Recurrence</td>
<td>CT (pre-tx)</td>
<td>I</td>
<td>79</td>
<td>Radiomics/Genomics</td>
</tr>
<tr>
<td>Depeursinge et al. 2015 [174]</td>
<td>Surgery</td>
<td>Recurrence</td>
<td>CT (pre-tx)</td>
<td>I</td>
<td>101</td>
<td>Reisz wavelets</td>
</tr>
<tr>
<td>Pyka et al. 2015 [179]</td>
<td>SABR</td>
<td>LR</td>
<td>PET/CT (pre-tx)</td>
<td>I</td>
<td>45</td>
<td>Texture</td>
</tr>
<tr>
<td>Wu et al. 2016 [180]</td>
<td>SABR</td>
<td>DM</td>
<td>PET/CT (pre-tx)</td>
<td>I</td>
<td>101</td>
<td>Radiomics</td>
</tr>
<tr>
<td>Aerts et al. 2014 [125]</td>
<td>RT &amp; ChemoRT</td>
<td>OS</td>
<td>CT (pre-tx)</td>
<td>I-IV</td>
<td>1019</td>
<td>Radiomics</td>
</tr>
<tr>
<td>Fried et al. 2014 [171]</td>
<td>ChemoRT</td>
<td>OS, LRC, FFDM</td>
<td>CT (pre-tx)</td>
<td>III</td>
<td>91</td>
<td>Texture</td>
</tr>
<tr>
<td>Coroller et al. 2015 [172]</td>
<td>ChemoRT</td>
<td>DM</td>
<td>CT (pre-tx)</td>
<td>II-III</td>
<td>182</td>
<td>Radiomics</td>
</tr>
<tr>
<td>Grove et al. 2015 [162]</td>
<td>Surgery</td>
<td>OS</td>
<td>CT (pre-tx)</td>
<td>I-IV</td>
<td>108</td>
<td>Tumour shape and heterogeneity</td>
</tr>
</tbody>
</table>
The use of radiomics to quantify the appearance of FDG-PET SUV changes can also be performed [179, 180]. This may be an important area of study to determine regional variations in SUV uptake pre- or post-treatment. A study by Pyka et al. evaluated the use of second-order GLCM texture features on pre-treatment FDG-PET images for predicting local recurrence following SABR for 45 patients with early-stage lung cancer [179]. They found that the entropy texture feature demonstrated an AUC of 0.87. Wu et al. analyzed 101 patients treated with SABR and found radiomic features on pre-treatment PET were associated with DM [180]. These techniques may be valuable for predicting outcomes in this patient population following SABR treatment. However, all of these studies need to ensure standardization of methodology as discretization of SUV values can have a major impact on the resultant texture features [181].

1.9 Thesis Hypothesis and Objectives

SABR has been shown to be to surgery for the treatment of patients with early-stage NSCLC. However, the presence of benign radiographic changes on post-SABR CT can occur and appear with similar shape and size to a recurring tumour. Being able to differentiate benign fibrotic changes from local tumour recurrence remains a major clinical challenge for timely response assessment following SABR. There is currently a great unmet clinical need to provide timely and accurate assessment of response following SABR for early stage NSCLC. Therefore, a reliable measure of recurrence on
CT imaging is critically needed, as the utilization of SABR is rapidly increasing and CT imaging remains the standard measure of imaging follow-up for these patients.

There are several advantages to the use of CT, rather than routine functional imaging, in assessment of response post-SABR. In contrast to FDG-PET imaging, CT is more accessible and inexpensive, does not rely on isotopes with short half-lives, and is already part of standard-of-care follow-up for patients who have received curative treatment for early-stage lung cancer, and who are eligible for salvage. Importantly, standardization of CT across centres is much less complex than standardization of PET/CT. Standard machine settings and reconstruction algorithms are widely available for CT imaging of the chest, increasing the generalizability of any follow-up recommendations. As such, new algorithms for early detection of recurrence based on standard-of-care CT imaging could be easily integrated into current clinical practice. However, novel imaging techniques must move beyond qualitative image analysis and simple RECIST measurements.

Quantitative image analysis allows for maximal information to be obtained from images already being performed in clinical practice, and can easily be translated into a useful clinical tool to aid in outcome prediction and treatment response assessment. Radiomics on pre-treatment imaging has demonstrated the potential to predict outcomes in all stages of lung cancer. To the best of our knowledge, there has been no previous work using radiomics to predict response using follow-up CT imaging. Physicians currently rely on qualitative image analysis to try to distinguish patients with recurrence from those with only benign injury, and typically rely on features of an enlarging mass. No prior studies have assessed physician performance or inter-physician variability and
reliability in detecting recurrence after SABR. We conjecture that post-SABR radiological changes can be quantified through the use of a radiomic software system to aid in response assessment after treatment.

To address the unmet clinical need to provide timely and accurate assessment of response following SABR for early stage NSCLC, the overarching objective of this thesis is to develop a radiomic software system to aid in treatment response assessment on follow-up CT after SABR. This objective will be evaluated by testing the following central hypothesis; a radiomic software system will outperform physicians in the early assessment of response post-SABR. To test the central hypothesis, this thesis has the following four aims:

1) To determine group differences in radiomic features between patients with recurrence and benign injury.

2) To develop a radiomics system to predict recurrence in individual patients.

3) To develop a semi-automated segmentation system to predict recurrence in individual patients.

4) To compare a radiomics decision support system to physician performance.
1.10 Thesis Outline

1.10.1 Chapter 2: Distinguishing radiation fibrosis from tumour recurrence after SABR for lung cancer: a quantitative analysis of CT density changes

This work was a hypothesis generating study to determine if quantitative image feature analysis could be used for distinguishing groups of patients with and without local recurrence following SABR. Quantitative image features were extracted from manually delineated regions of common post-SABR changes; consolidation and GGO. This will allow us to assess the potential for quantitative image appearance features to provide additional information over traditional size measurements. This could allow for quantitative appearance features to be used in a decision support tool for aiding in response assessment post-SABR. We hypothesized that there would be quantitative differences in appearance features between recurrence and benign fibrosis patient groups.

1.10.2 Chapter 3: Early prediction of tumour recurrence based on CT texture changes after SABR for lung cancer

The objective of this work was to evaluate quantitative image feature analysis and machine learning techniques for predicting local recurrence in individual patients following SABR. A more comprehensive set of first- and second-order image features was extracted from regions of post-SABR changes and machine learning was performed to evaluate classification performance. We wanted to work towards the goal of understanding if second-order quantitative image features can predict local recurrence in individual patients within 6 months of SABR treatment. We also aimed to assess the impact of variations in manual segmentation boundaries on classification performance.
and compare the quantitative appearance features to traditional measure of response post-SABR. We hypothesized that quantitative appearance features can predict recurrence within 6 months of SABR with accuracies greater than 75%.

1.10.3 Chapter 4: Imaging texture analysis for automated prediction of lung cancer recurrence after stereotactic radiotherapy

The objective of this work was to develop and evaluate a semi-automated graph cuts based approach to segment and sample regions of post-SABR radiological changes. The work in Chapters 2 and 3 relied on manual delineations of post-SABR radiological changes on CT which are time consuming and subject to operator variability. We look to determine if semi-automated segmentation methods can decrease segmentation time and produce consistent segmentations between operators. In addition to developing a semi-automated segmentation approach, we wanted to determine if this system can be used with the radiomic decision support system to predict local recurrence following SABR. We hypothesized that a semi-automated segmentation algorithm could demonstrate consistent segmentations and prediction accuracies between operators, with non-inferiority to manual segmentations.

1.10.4 Chapter 5: Detection of local cancer recurrence after SABR for lung cancer: physician performance versus radiomic assessment

The objective of this work was to perform an observer study to assess how well physicians can distinguish between benign fibrosis and local recurrence on follow-up CT after SABR. This would allow for a basis of comparison for our radiomics system and allow us to move towards the goal of understanding how physicians perform in response
assessment post-SABR. To the best of our knowledge, there has been no study to date assessing a physician’s performance in distinguishing fibrosis from recurrence following SABR for lung cancer. An additional objective of this work was to develop and evaluate a multi-feature radiomic signature to predict local recurrence following SABR. In this chapter we compared our radiomic signature to physician performance for the early prediction of recurrence within 6 months of treatment. We hypothesized that a radiomic software system will outperform physicians in the early assessment of response post-SABR.

1.11 References


110. S. Feigenberg, J. Yu and T. Eade, "FDG PET response by 3 months following stereotactic body radiotherapy for non-small cell lung cancer may be an early surrogate of local failure [abstract]," Int J Radiat Oncol Biol Phys 69 (3 Suppl.1), S479-S480 (2007).


Chapter 2

Towards the goal of understanding the quantitative image features between benign radiation induced lung injury and local recurrence, we evaluated quantitative image feature analysis for distinguishing groups of patients with and without local recurrence following SABR.

The contents of this chapter were previously published in the journal Acta Oncologica: SA Mattonen, DA Palma, CJ Haasbeek, S Senan, and AD Ward. Distinguishing radiation fibrosis from tumour recurrence after stereotactic ablative radiotherapy (SABR) for lung cancer: A quantitative analysis of CT density changes. Acta Oncol 2013; 52:910–918; http://dx.doi.org/10.3109/0284186X.2012.731525. Permission to reproduce this article was granted by Taylor & Francis Group and is provided in Appendix A.3.

2 DISTINGUISHING RADIATION FIBROSIS FROM TUMOUR RECURRENCE AFTER SABR FOR LUNG CANCER: A QUANTITATIVE ANALYSIS OF CT DENSITY CHANGES

2.1 Introduction

Stereotactic ablative radiotherapy (SABR), also known as stereotactic body radiotherapy (SBRT), employs image guidance and precise treatment delivery to deliver a high dose of radiation to tumours while sparing surrounding normal tissue [1]. Treatments are typically delivered with high doses per fraction, over 3-8 fractions [2]. SABR is becoming a standard treatment option for stage I non-small cell lung cancer (NSCLC) patients who are medically inoperable or refuse surgery [2, 3]. Following lung radiotherapy, radiation-induced lung injury (RILI) can occur in the acute phase (within 6 months) as radiation pneumonitis and in the late phase (after 6 months) as fibrosis [4]. With SABR, the incidence of acute- and late-onset RILI is high: 54-79% of patients develop acute benign CT changes, and 80-100% of patients develop late changes [5].
With older radiotherapy techniques, benign computed tomography (CT) changes typically follow the edges of the treatment fields, consisting of straight borders and allowing for easier diagnosis of RILI [4]. Due to the high doses delivered and steep dose gradients with SABR, as well as the 3D conformity of the treated region to the morphology of the target, benign CT changes are common and can mimic tumour recurrence, especially when they develop as mass-like patterns [5, 6]. Therefore, distinguishing between benign fibrosis and recurrence becomes of major importance in identifying patients who are candidates for salvage treatment [7]. Misclassification of injury as recurrence can cause patients to undergo unnecessary investigations and interventions for no oncologic benefit. Conversely, misclassification of recurrence as injury can result in a missed opportunity for salvage [6].

\(^{18}\text{F-fluorodeoxyglucose positron emission tomography (FDG-PET) imaging has been investigated to detect recurrence, but has several limitations: hypermetabolic activity has been observed in benign SABR-treated lesions [8], optimal SUV cut-offs have not been defined, and repeated PET scanning of all SABR patients would be costly. A biopsy can be considered for a definite diagnosis of recurrence; however, high risks of complications such as pneumonothorax (\(\sim\)20%) and hemoptysis (5-10%) prevent it from being a first line tool for differentiating fibrosis from recurrence [9]. As a result, better methods are needed to aid in distinguishing RILI from recurrence. A reliable measure of recurrence on CT imaging would be a valuable clinical tool, as the utilization of SABR for early stage lung cancer is increasing [10], and CT scanning is the standard measure of imaging follow-up for SABR patients [11].
The goal of this study is to measure the utility of CT image feature analysis in differentiating RILI from recurrence, when compared to Response Evaluation Criteria in Solid Tumours (RECIST) measurements, which is the standard method used for response assessment in trials of oncology [12]. Specifically, this study compares the imaging characteristics of CT density, texture and 3D volume of contoured regions of interest: consolidative regions (increased density with no visibility of vessels) and ground-glass opacity (GGO: an increase in normal lung parenchyma density but with visibility of vessels). Analysis was completed on planning and follow-up CT images, and determines which measures may provide the earliest differentiation of local recurrence from injury.

2.2 Methods and Materials

2.2.1 Patient Selection and Treatment

A total of 22 patients with 24 tumours treated with SABR for Stage I NSCLC at the VU University Medical Center between February 2004 and February 2010 were selected for this study. Patients without recurrence were selected from a previous study [13] based on presence of CT findings of moderate/severe fibrosis that were ultimately found to be benign. Of the 24 lesions, 13 developed only benign RILI, and 11 developed recurrence. Of the latter, pathological evidence of recurrence was available in 8 of 11 lesions; the other 3 were deemed to be recurrences based on sequential imaging findings and multidisciplinary group consensus, as biopsy was not feasible. One patient had 2 treated lesions in close proximity; these were considered as a single lesion for analysis as the post-SABR CT changes spanned the location of both lesions.
All patients were treated with SABR using a risk-adapted fractionation scheme as previously described [13-15]. Briefly, patients were treated with one of three fractionation regimens: 60 Gy in 3 fractions for T1 tumours surrounded by lung parenchyma; 60 Gy in 5 fractions for T2 tumours or those in broad contact with the chest wall; and 60 Gy in 8 fractions for tumours within 2 cm of the mediastinum or close to the brachial plexus. For 3D-conformal treatments, done using 9-11 fixed beams, planning was completed using BrainLab software (Brainscan version 5.2; BrainLab Inc., Feldkirchen, Germany) and a pencil beam dose calculation algorithm. For those patients treated with volumetric modulated arc therapy (VMAT) using RapidArc (Varian Medical Systems, Palo Alto, USA), the anisotropic analytical algorithm (AAA) was used for dose calculations. With the latter, the prescribed doses for the 3- and 5- fraction regimens were 54 Gy and 55 Gy respectively, due to differences in dose calculation with the more advanced algorithm. Follow-up was routinely conducted with a diagnostic CT scan approximately 3, 6, and 12 months post treatment and every 6-12 months thereafter.

### Contours

All pre-treatment and follow-up scans were manually contoured on every axial slice using a lung window setting (window width of 1500 HU and window level of -600 HU). For tumours or fibrosis abutting the mediastinum, a mediastinal window setting (window width of 350 HU and window level of 40 HU) was also used. All contours were completed using ITK-SNAP [16] (Version 2.2.0). Two regions were contoured in the follow-up scans based on the patterns of post-SABR CT changes which have been well described [13, 15, 17]. The first, containing consolidative changes around the treatment site, was defined as a region of increased density with no visibility of vessels. The
second, containing ground glass opacities (GGO), was defined as an increase in normal lung parenchyma density but with visibility of vessels [18]. Figure 2-1 shows an example of the contours created for a single subject throughout the course of follow-up post-SABR. All contours were completed by a single investigator (SAM) and subsequently edited and approved by a thoracic radiation oncologist (DAP). Further details pertaining to image acquisition and analysis are outlined in Appendix B.

Inter- and intra-observer variability has been reported in contouring of lung tumours [19]. To simulate the effect of contouring variability, we implemented computer software to expand and contract the contour borders by 1 mm in the axial plane, and determined the effect of the expansion and contraction on the mean differences measured between patients with RILI and recurrence. At the earliest significant time point for both the consolidative and GGO regions, we expanded and contracted the borders by 1 mm and recalculated our results.

![Figure 2-1](image)

**Figure 2-1:** Manual contours throughout a course of follow-up for a patient whose cancer recurred. The solid lines represent consolidative changes, and the dashed lines represent ground-glass opacity (GGO).
2.2.3 Data Analysis

Four measures were calculated for the contoured regions: (1) mean 3D volume, (2) mean CT density, (3) standard deviation of CT density, and (4) mean RECIST measurements. RECIST measurements were taken for only the solid consolidative areas according to RECIST 1.1. MATLAB 7.13 (The MathWorks Inc., Natick, MA, USA) was used in the calculation of the 3D volume and CT densities of the contoured regions. The standard deviation (SD) of CT density was used as a basic texture metric, where larger standard deviations correspond to more variegated image textures within the contoured regions. Figure 2-2 shows an example of this image texture measure for three sample image regions containing increasingly variegated texture, with the corresponding standard deviations shown in HU.

Figure 2-2: The standard deviation (SD) of CT density was used as a basic texture measure. Three examples from GGO regions of patients in this study, showing the differences in texture with the varying SDs: a) 50.1 HU, b) 85.8 HU, c) 123.4 HU

For each lesion, the measures were analyzed cumulatively throughout the course of follow-up up to a specified time point. This resulted in a mean RECIST size, mean 3D volume, mean CT density, and standard deviation of CT density for each lesion over each specified follow-up time interval. The time points we used for analysis were every 3
months up to 2 years, then yearly thereafter. These time points were more frequent than the regular follow-up scan intervals as there is occasionally some variability in CT scan timing due to individual patient scheduling and potentially due to clinical indications. The end date of treatment was considered day 0 of follow-up. Differences between groups were assessed using an independent samples t-test with unequal variances in MATLAB 7.13 (The MathWorks Inc., Natick, MA, USA). All statistical tests were two-sided with $p \leq 0.05$ indicative of statistical significance.

2.3 Results

A total of 136 CT scans were reviewed with a median imaging follow-up of 26 months (range: 6 - 44 months). The mean number of scans per patient was 6 (range 2-8); patients with recurrence had a mean of 4 follow-up scans, and patients without recurrence had a mean of 6 follow-up scans. Baseline patient and treatment variables are shown in Table 2-1. We observed that over 50% of tumours were found in the upper lobe, which is consistent with epidemiologic studies of lung cancer [20]. Three patients developed clinical grade 3 pneumonitis, and no grade 4 or 5 thoracic clinical toxicity was observed in this patient cohort. All patients without recurrence had a minimum follow-up of 2 years.
Table 2-1: Baseline patient and treatment variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total (n=22 patients, with 24 tumours)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median (range) or Number (%)</td>
</tr>
<tr>
<td>Age (n=22)</td>
<td>69 (49-84)</td>
</tr>
<tr>
<td>Gender (n=22)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>14 (63.6%)</td>
</tr>
<tr>
<td>Female</td>
<td>8 (36.4%)</td>
</tr>
<tr>
<td>Charlson Score* (n=22)</td>
<td>5 (1-10)</td>
</tr>
<tr>
<td>Planning target volume (n=24)</td>
<td>20.4 cc (4.5-144.4 cc)</td>
</tr>
<tr>
<td>Fractionation schedule (n=24)</td>
<td></td>
</tr>
<tr>
<td>3 x 20 Gy</td>
<td>14 (58.3%)</td>
</tr>
<tr>
<td>5 x 11 Gy</td>
<td>1 (4.2%)</td>
</tr>
<tr>
<td>5 x 12 Gy</td>
<td>4 (16.7%)</td>
</tr>
<tr>
<td>8 x 7.5 Gy</td>
<td>5 (20.8%)</td>
</tr>
<tr>
<td>Treatment type (n=24)</td>
<td></td>
</tr>
<tr>
<td>Fixed beam</td>
<td>21 (87.5%)</td>
</tr>
<tr>
<td>RapidArc</td>
<td>3 (12.5%)</td>
</tr>
<tr>
<td>Tumour location (n=24)</td>
<td></td>
</tr>
<tr>
<td>Left upper lobe</td>
<td>4 (16.7%)</td>
</tr>
<tr>
<td>Left lower lobe</td>
<td>2 (8.3%)</td>
</tr>
<tr>
<td>Right upper lobe</td>
<td>10 (41.7%)</td>
</tr>
<tr>
<td>Right middle lobe</td>
<td>3 (12.5%)</td>
</tr>
<tr>
<td>Right lower lobe</td>
<td>5 (20.8%)</td>
</tr>
</tbody>
</table>

*age adjusted Charlson, current lung cancer not scored as comorbidity.
2.3.1 Pre-treatment lesions

Table 2-2 shows the baseline pre-treatment imaging characteristics for patients with recurrence and RILI. No significant differences were observed at baseline in mean CT density, volume, or RECIST size, comparing lesions of patients whom would later develop recurrence versus those with RILI. Since GGO was not contoured on the pre-treatment scan, the standard deviation of CT density was not included in the aforementioned table.

**Table 2-2: Baseline pre-treatment imaging characteristics.**

<table>
<thead>
<tr>
<th></th>
<th>Recurrence Mean ± 95% CI</th>
<th>Benign Fibrosis Mean ± 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-treatment CT Density (p=.434)</td>
<td>-104.8 ± 34.40 HU</td>
<td>-124.6 ± 34.43 HU</td>
</tr>
<tr>
<td>Pre-treatment RECIST (p=.052)</td>
<td>3.55 ± 1.01 cm</td>
<td>2.36 ± 0.43 cm</td>
</tr>
<tr>
<td>Pre-treatment 3D Volume (p=.096)</td>
<td>22.29 ± 16.27 cc</td>
<td>6.67 ± 4.66 cc</td>
</tr>
</tbody>
</table>

2.3.2 Post-SABR Measures of Size

A significant difference in RECIST measurements was detected between patients with recurrence vs. those with RILI as early as 15 months post-treatment (p=0.028). Patients with recurrence had a mean cumulative RECIST [± 95% CI] of the consolidative changes at 15 months of 4.34 ± 1.13 cm versus 2.63 ± 0.84 cm for patients with benign RILI.
Patients with recurrence had significantly larger solid consolidative changes as measured with 3D volume, detected as early as 15 months post-treatment compared to patients with RILI (mean at 15 months [± 95% CI] of 30.1 ± 19.3 cc vs. 5.1 ± 3.6 cc, respectively; p=0.030). No significant difference was observed in the 3D volumes of the GGO regions. Figure 2-3 shows the significant measures of size (both RECIST and 3D volume measurements) of the consolidative regions for patients with recurrence and RILI, during the course of follow-up post-SABR.
Figure 2-3: Cumulative size measures of the consolidative regions throughout follow-up post-SABR; all values are the mean ± 95% CI. a) RECIST and b) 3D volume. * Indicates statistical significance at p ≤ 0.05.
2.3.3 Post-SABR Appearance Measures

Patients with recurrence had significantly denser mean [± 95% CI] solid consolidative changes of -96.4 ± 32.7 HU versus -143.2 ± 28.4 HU for patients with RILI, and this was detected as early as 9 months post-treatment (p=0.046). No significant difference was observed in the CT density of the GGO at any point in follow-up.

Significantly increased variability of CT densities in the GGO areas was detected at 9 months post-treatment (p=0.0078). Patients with recurrence had a standard deviation of CT density at 9 months [± 95% CI] of 210.6 ± 14.5 HU vs. 175.1 ± 18.7 HU for patients with RILI. This is indicative of larger variability in the HU of the GGO, or more variegated texture, in patients with recurrence. No significant difference was observed in the variation of CT densities within the consolidative regions. Figure 2-4 and Figure 2-5 show the cumulative appearance measures (CT density and standard deviation of CT density), in the GGO and consolidative regions respectively, during post-SABR follow-up for patients with recurrence and RILI.
Figure 2-4: Cumulative appearance measures of the ground glass opacity regions throughout follow-up post-SABR; a) Mean [± 95% CI] CT density, and b) standard deviation [± 95% CI] of CT density.

* Indicates statistical significance at p ≤ 0.05
Figure 2-5: Cumulative appearance measures of the consolidative regions throughout follow-up post-SABR; a) Mean [± 95% CI] CT density, and b) standard deviation [± 95% CI] of CT density.

* Indicates statistical significance at p ≤ 0.05
2.3.4 Contouring Variability

In the consolidative regions at 9 months, we expanded and contracted the borders by 1 mm and recalculated our results. In all cases, including the original results, the differences between the two groups remained between 36-59 HU, with the patients with recurrence having higher HU density than patients with only RILI: [1 mm expansion: RILI = -220.02 ± 35.70 HU, recurrence = -161.67 ± 36.26 HU; 1 mm contraction: RILI = -95.43 ± 28.48 HU, recurrence = -59.58 ± 33.06 HU].

Similarly, for the GGO texture measure at 9 months, patients with recurrence had a higher standard deviation of CT density compared to those with RILI, and the difference between the two groups remained between 35-39 HU: [1 mm expansion: RILI = 194.77 ± 14.99 HU, recurrence = 229.97 ± 10.85 HU; 1 mm contraction: RILI = 167.86± 21.71 HU, recurrence = 206.60 ± 17.16 HU].

2.4 Discussion

CT density changes are common after SABR and distinguishing recurrence from RILI is becoming increasingly important. As SABR utilization increases, ambiguous CT findings will become a more common clinical problem. A technique for early and accurate diagnosis of post-SABR recurrence could allow for early salvage of recurrence, and avoid unnecessary imaging and intervention in patients with RILI only. This study suggests that changes in the cumulative mean density of consolidative regions and textural analysis of the GGO have the potential to distinguish RILI from cancer recurrence as early as 9 months post-SABR, compared to 3D volume and RECIST at 15 months. For RECIST of the consolidation and texture analysis of the GGO, we
acknowledge the possibility that they may be inferior predictors as there does not appear to be an increase in separation between the groups over time. To our knowledge, this is the largest series of recurrences analysed for post-SABR radiological changes in the literature.

We observed a significant difference in mean HU for the consolidative regions, but not in the GGO regions, and conversely we observed a significant difference in the standard deviation of HU in the GGO regions, but not in the consolidative regions. A more detailed investigation is required in order to more fully understand the reasons behind this observation. Since the consolidative regions are relatively opaque and homogeneous in appearance on CT imaging, the standard deviation of HU in these regions may be a less sensitive measure of difference between the groups, and the significant difference in mean HU could potentially be due to changes in tissue composition of the recurrent tumour (such as microvasculature and soft tissue) compared to fibrotic tissue within radiation-induced lung injury.

The GGO regions contained a relatively greater amount of variation in the CT densities as some areas have lower density lung tissue, and others have higher density vessels; this may have rendered the standard deviation of HU measure a relatively more sensitive measure of difference between the groups in the GGO regions. We surmise that RILI results in a more uniform GGO density within these regions, whereas recurrences seemed to have a more variegated appearance potentially representing tumour nodularity or increased vascularity in areas of tumour recurrence. Further work is required to gain an understanding of the exact reasons behind the association of our significant measures with the consolidation or GGO regions for patients with recurrence and RILI.
The findings of this study are consistent with previously published literature but extend these previous findings in several important ways. Previous studies have shown that an enlarging opaque region after 12 months is indicative of recurrence [21, 22], consistent with the RECIST findings reported herein. We hypothesize that the decrease in RECIST and volume of the consolidative regions for the recurrence groups at 3 months is indicative of a partial response to treatment, and following this time point the rise we see may possibly be a sign of progressive disease. However, an enlarging mass may not be specific for recurrence: based on biopsy and further imaging, only a few patients with enlarging masses are deemed to ultimately have recurrence [6]. Most previous studies of radiological features of post-SABR recurrence are limited due to small sample size (of ≤5 recurrences) and mainly focused on the type and qualitative appearance of these changes [6, 21, 22]. Our study extends these findings by quantifying changes in density, texture, and 3D volume for patients, and suggests that these measurements may allow for the earlier detection of recurrence.

A systematic review by Huang et al. identified several predictive factors of recurrence [5]. Since the primary imaging modality for follow-up post-SABR is CT, RECIST is traditionally used to determine a patient’s response to treatment. An enlarging mass with 20% increase in size, or an absolute increase in size of at least 5 mm from baseline, as described by RECIST 1.1, is suspicious for recurrence and warrants FDG-PET for further investigation. Other high-risk features on CT include an enlarging mass-like lesion, opacity enlargement after 12 months, bulging margin, disappearance of air bronchograms, linear margin disappearance, ipsilateral pleural effusion, or lymph node enlargement [5]. In terms of metabolic findings of recurrence, a SUVmax ≥ 5 using FDG-
PET is highly suspicious of recurrence. In these cases, it is suggested a biopsy and/or resection should occur if feasible. Validation of the findings presented herein may improve this algorithm for imaging follow-up after SABR.

The conclusions of this study must be considered in the context of its strengths and limitations. Although the number of recurrent lesions analyzed is more than in previous studies, and the length of follow-up is long, the sample size remains small, which may limit the power to detect small differences between groups. The patients included are not completely reflective of the overall SABR population, in that both the proportion of patients with recurrence, and the severity of benign RILI are higher than in a general SABR population. This sample was selected in order to determine if image feature analysis could differentiate recurrence from difficult RILI cases, but these findings require validation in a large, separate dataset. Inter- and intra-observer variability can exist in contouring of lung tumours [19], and further research is required to refine these metrics so they are invariant to individual contouring practices. As a result, it appears that even in the presence of contouring variability, the differences between the RILI and recurrence group appear consistent. Furthermore, data analysis was completed through cumulative analysis of measures up to a time point (0-3, 0-6, 0-9 months etc.), rather than using smaller time ranges (0-3, 3-6, 6-9 months etc.). Future studies will evaluate the relative benefits of using a cumulative analysis vs. using smaller time ranges. Furthermore, lung density can vary with regional blood flow, respiration, as well as co-morbid illnesses such as COPD [23, 24], and these factors were not considered in this study. Further work must be completed on the effect of any co-morbidities on the measures analysed in this study.
In conclusion, our study suggests that the use of CT density of consolidative changes as well as basic textural analysis of GGO may allow for early differentiation between RILI and recurrence. With further validation of our results on a larger sample size, and more detailed analysis of the features and changes observed throughout the course of SABR follow-up, there is the potential for an earlier detection of recurrence compared with traditional measures. This could potentially allow for earlier salvage of patients with recurrence, and result in fewer investigations for patients exhibiting only benign RILI.

2.5 References


Chapter 3

Towards the goal of understanding if quantitative image features can predict local recurrence in individual patients, we evaluated quantitative image feature analysis plus machine learning techniques for predicting local recurrence in individual patients following SABR.

The contents of this chapter were previously published in the journal Medical Physics: SA Mattonen, DA Palma, CJ Haasbeek, S Senan, and AD Ward. Early prediction of tumor recurrence based on CT texture changes after stereotactic ablative radiotherapy (SABR) for lung cancer. Med Phys 2014; 41(3): 033502. Permission to reproduce this article was granted by the American Association of Physicists in Medicine and is provided in Appendix A.4.

3 EARLY PREDICTION OF TUMOUR RECURRENCE BASED ON CT TEXTURE CHANGES AFTER SABR FOR LUNG CANCER

3.1 Introduction

Stereotactic ablative radiotherapy (SABR) is now a standard treatment option for patients with early stage non-small cell lung cancer (NSCLC) who are medically inoperable or refuse surgery [1, 2]. SABR uses advanced treatment planning and delivery to treat the tumour to a high dose, while sparing surrounding normal tissue. Multiple collimated radiation beams are used to achieve a dose distribution highly conformal to the shape of the tumour with steep dose gradients at the tumour boundary. SABR treatments are typically delivered with high doses per fraction with relatively fewer fractions in the overall treatment (e.g. 7.5–18 Gy/fraction over 3–8 fractions for SABR, compared to approximately 2 Gy/fraction over 20–30 fractions for conventional therapy) [3].
Currently, the standard imaging modality recommended and clinically used for post-SABR follow up is computed tomography (CT) [2]. During follow-up assessment, a key clinical decision is whether to provide further, possibly more invasive intervention (e.g. surgery, chemotherapy) to treat or remove recurrent/residual disease. This decision rests on the ability to assess the success of the SABR treatment; i.e. to determine whether the patient’s cancer will recur. Since recurrent lung cancer typically progresses quickly, a decision to proceed with further intervention is most valuable if made early, since delayed detection of recurrence may reduce the options for salvage therapies. This decision is complicated by the fact that following radiotherapy to the lung, radiation-induced lung injury (RILI) can occur as radiation pneumonitis and radiation fibrosis, which appear as an increase in lung density on CT [4, 5]. Following treatment with SABR, RILI can have similar size and morphology to a recurrent tumour [6, 7]. Several studies have examined the radiologic appearance of recurrence on follow-up CT post-SABR, and suggest that an enlarging opacity after 12 months is most suggestive of recurrence; however, this criterion can only be met one year after treatment [8, 9]. A systematic review of the literature also suggests that other imaging features, such as a bulging margin and disappearance of air bronchograms, are also suggestive of recurrence [10, 11].

Metabolic imaging modalities such as 18-fluorodeoxyglucose (FDG) positron emission tomography (PET) imaging have been investigated for measuring treatment response post-SABR. Standardized uptake values (SUVs) greater than 5, or greater than the pre-treatment value, have been shown to be suggestive of recurrence; however, optimal SUV thresholds have not been determined [11]. Adding to the challenge of
interpreting FDG-PET images in the post-SABR context, hypermetabolic activity has been observed in benign fibrotic tissues years following SABR, likely due to an inflammatory response [12].

A means for predicting recurrence within 6 months of treatment based on routine CT imaging would permit timely intervention for recurrence, which typically manifests after 1 year [3], and would be ideally suited to tailoring clinical management. Image texture analysis has been used for computer-aided diagnosis on lung CT, and second-order texture statistics based on grey-level co-occurrence matrices (GLCMs) have been shown to quantify lung abnormalities [13, 14]. Our previous work presented in Chapter 2, demonstrated that a first-order texture feature [standard deviation (SD) of CT density] within the ground-glass opacity (GGO) regions, as well as the density of the consolidative regions, could statistically significantly differentiate recurrence and RILI patient groups at 9 months post-SABR [15]. Our preliminary analysis showed that a linear classifier could predict recurrence in individual patients at 9 months post-SABR with error of 26% using first-order texture of the GGO [16]. To the best of our knowledge, no previous work has demonstrated the ability to predict eventual cancer recurrence based on images acquired within 6 months of SABR treatment.

Based on our observations of the utility of quantitative appearance measures of the GGO in Chapter 2 [15], the primary objective of this study was to measure the accuracy of second-order GGO texture features on CT images acquired within 6 months of SABR treatment for prediction of recurrence, and to compare this with the accuracies obtained using first-order appearance features and measures of size of the consolidation regions. We analyzed texture features within GGO regions known to undergo
radiographic changes post-SABR [17]. Based on our observation that contouring of the consolidation regions is relatively straightforward in a clinical context, but GGO contouring is highly time-consuming and requires more careful judgment, our secondary objective was to measure the sensitivity of classification performance to perturbations of the GGO boundaries. This study is intended to generate hypotheses regarding the relative merit of GGO appearance features for predicting recurrence for further testing on a larger data set.

3.2 Materials and Methods

3.2.1 Patient Selection and Imaging

A group of 22 patients with T1 or T2 N0 non-small cell lung cancer was selected for this study, all of whom were treated with SABR at the VU University Medical Center in the Netherlands. These patients had a total of 24 tumours; 13 developed benign RILI, and 11 developed recurrence. Pathological evidence of recurrence was available in 8 of the 11 tumour cases. Biopsy was not feasible in the remaining 3, and they were deemed to be recurrences based on sequential imaging findings, multidisciplinary group discussion, and eventual clinical outcome. All patients with benign RILI in this study had at least 2 years of imaging follow-up and were selected based on moderate to severe radiological findings [17], thus presenting a challenge to an algorithm designed to distinguish the substantial CT changes in these RILI cases from true recurrences. Patient characteristics and treatment variables are provided in Table 3-1.
Table 3-1: Baseline patient characteristics and treatment variables.

<table>
<thead>
<tr>
<th></th>
<th>Total (n=22 patients with 24 tumours)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median (range) or Number (%)</td>
</tr>
<tr>
<td><strong>Age (n=22)</strong></td>
<td>69 (49-84)</td>
</tr>
<tr>
<td><strong>Gender (n=22)</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>14 (63.6%)</td>
</tr>
<tr>
<td>Female</td>
<td>8 (36.4%)</td>
</tr>
<tr>
<td>*<em>Charlson Score</em> (n=22)**</td>
<td>5 (1-10)</td>
</tr>
<tr>
<td><strong>Tumour location (n=24)</strong></td>
<td></td>
</tr>
<tr>
<td>Left upper lobe</td>
<td>4 (16.7%)</td>
</tr>
<tr>
<td>Left lower lobe</td>
<td>2 (8.3%)</td>
</tr>
<tr>
<td>Right upper lobe</td>
<td>10 (41.7%)</td>
</tr>
<tr>
<td>Right middle lobe</td>
<td>3 (12.5%)</td>
</tr>
<tr>
<td><strong>Planning target volume (n=24)</strong></td>
<td>20.4 cc (4.5-144.4 cc)</td>
</tr>
<tr>
<td><strong>Fractionation schedule (n=24)</strong></td>
<td></td>
</tr>
<tr>
<td>3 x 20 Gy</td>
<td>14 (58.3%)</td>
</tr>
<tr>
<td>5 x 11 Gy</td>
<td>1 (4.2%)</td>
</tr>
<tr>
<td>5 x 12 Gy</td>
<td>4 (16.7%)</td>
</tr>
<tr>
<td>8 x 7.5 Gy</td>
<td>5 (20.8%)</td>
</tr>
<tr>
<td><strong>Treatment type (n=24)</strong></td>
<td></td>
</tr>
<tr>
<td>Fixed beam</td>
<td>21 (87.5%)</td>
</tr>
<tr>
<td>RapidArc</td>
<td>3 (12.5%)</td>
</tr>
</tbody>
</table>

*Age adjusted Charlson, current lung cancer not scored as comorbidity.

All of the post-SABR follow-up CT images were acquired at the VU Medical Center in Amsterdam, Netherlands. Diagnostic CT scans were performed on one of three scanners: Siemens Volume Zoom 4-slice, Siemens Sensations 64-slice (Siemens Nederland N.V., Den Haag, Netherlands), or Philips Brilliance iCT 256-slice (Royal
Philips Electronics, Inc., Amsterdam, Netherlands). Standard machine settings were 120 kVp, 100 mAs, spiral acquisition, 0.5 second rotation time, 2.5–5 mm slice thickness, and images were acquired at inspiratory breath hold. Intravenous contrast (Ultravist-300; Bayer Pharma AG, Berlin, Germany) was administered at a volume of 70 cc and a delay of 25 seconds. Post-SABR diagnostic images were scheduled to be taken at approximately 3, 6, and 12 months following treatment, then every 6 to 12 months thereafter. However, due to practical considerations, the actual timing of follow-up scans does not conform exactly to the above schedule. Figure 3-1 shows the recorded timing of follow-up scans for the patients in our study, where the last date of treatment was considered day 0 of follow-up. We observed clustering of the first and second follow-up scans within the 2–5 and 5–8 month intervals post-treatment, respectively, and these were therefore the two time intervals used for this study. A total of 46 scans were analyzed within these two time intervals.

**Figure 3-1:** Distribution of post-SABR CT scan time points for the patients in this study. Patients marked as recurrences on this plot experienced eventual cancer recurrence, diagnosed later than the imaging time points indicated on this plot.
3.2.2 Region of Interest Delineation

Manual segmentations of two regions of interest were obtained from the study presented in Chapter 2 [15]. Two three-dimensional (3D) regions of common post-SABR radiographic changes were manually contoured using the ITK-SNAP [18] (Version 2.2.0) software package: consolidative and GGO regions. Segmentations were completed on axial image slices using a lung window setting (window width of 1500 Hounsfield units [HU] and window level of -600 HU). A mediastinal window (window width of 350 HU and window level of 40 HU) was also used for delineation of any structures abutting the mediastinum. All segmentations were completed by a single graduate student, reviewed by a thoracic radiation oncologist with expertise in lung SABR, and approved after any necessary editing. Consolidative regions were defined as having increased density with respect to the surrounding region, with no visibility of blood vessels within. GGO regions were defined by an increase in the normal lung parenchyma density with respect to the surrounding region, with potential visibility of blood vessels within [19]. Figure 3-2 illustrates the consolidative and GGO delineations for a patient with benign RILI and one with recurrence; note in this example the similarity of changes in the consolidative regions throughout the course of follow-up for the two patients, as well as qualitative differences between the recurrence and the RILI in the image textural qualities, particularly in the GGO regions.
3.2.3 Measures of Size

Two measures of size were taken in the consolidative region: the one-dimensional (1D) longest diameter as observed on any axial slice (henceforth referred to as the *longest axial diameter*) and the 3D volume. The largest axial diameter was used to make a *Response Evaluation Criteria in Solid Tumours* (RECIST) assessment for each consolidative region, according to RECIST 1.1[20]. All images were imported into ClearCanvas Workstation 2.0 (ClearCanvas Inc., Toronto, Canada) in anonymized form.
and the longest axial diameter measurements were taken using the ClearCanvas Workstation ruler. The 3D volumes of the consolidative regions were calculated using MATLAB 7.13 (The MathWorks Inc., Natick, MA, USA) by multiplying the number of voxels contained within each region by the voxel volume.

3.2.4 Measures of Imaging Appearance

Two first-order appearance measures were calculated within the GGO regions: mean CT density, and standard deviation of CT density (the latter as a measure of image texture, with higher standard deviation values indicative of more variegated textures). MATLAB 7.13 was used to calculate these first-order measures.

Seven second-order appearance measures were calculated within the GGO regions. Second-order image texture features measure, in various ways, the intensity relationships between pairs of voxels in the image; typically, voxels forming a pair are spatial neighbours. The first step in calculating such measures is to compute a gray-level co-occurrence matrix (GLCM) [21]. This is a two-dimensional square matrix \( g \) where the rows and columns correspond to observable gray levels (or gray level ranges) within the images, and where each matrix element \( g(i,j) \) contains a non-negative integer indicating the number of voxel pairs whose elements have gray levels \( i \) and \( j \). Our chosen seven texture features were calculated based on the Conners, Trivedi, and Harlow feature set [22-24]: energy, entropy, correlation, inverse difference moment (IDM), inertia, cluster shade, and cluster prominence. Their definitions and equations are provided in Table 3-2, where the weighted pixel average \( \mu = \sum_{i,j} i \cdot g(i,j) = \sum_{i,j} j \cdot g(i,j) \) (due to symmetry of \( g \)), and the weighted pixel variance \( \sigma = \sum_{i,j}(i - \mu)^2 \cdot g(i,j) = \sum_{i,j}(j - \mu)^2 \cdot g(i,j) \)
(due to symmetry of $g$). These second-order texture features were calculated using the Insight Segmentation and Registration Toolkit (ITK) 4.3.1, an open source image processing software library available at www.itk.org [25].

Computation of second-order texture features requires the configuration of the GLCM histogram bins and the spatial directions used to establish pairings of neighbouring voxels. Histogram distributions of the CT densities within the GGO regions were analyzed to determine the appropriate number and density ranges of the bins for the GLCMs. Within the GGO, densities ranged from -1000 HU to 200 HU. The number of density bins was set accordingly to 60 bins, yielding 20 HU bin widths for analysis. In CT images there are thousands of possible greyscale values and therefore the allocation of a GLCM bin to every possible greyscale value will result in a very large GLCM. This can cause texture feature values to become very sensitive to the number of voxels in the sample. Since we are analyzing a small region of interest, we wanted to reduce the number of intensities per GLCM bin for analysis. Based on an inspection of one patient image, a bin width of 20 HU was chosen as a balance that both sufficiently quantizes the intensities while still having sufficient grey levels for discrimination of intensity differences in the image. This bin width parameter was not precisely tuned to optimize the results; improved results may be possible via optimization of this parameter in the context of a larger samples size. Neighbouring voxels were paired in four spatial directions within the 2D axial image planes $[(-1, 0), (-1, -1), (0, -1), (1, -1)]$. All second-order texture measures were computed based on GLCMs calculated using each of these four directions, and were averaged over all directions. When calculating a second-order texture features a single direction for analysis results in a directional dependency of that
feature. Therefore, calculating over multiple directions mitigates dependency on the spatial orientation of the image (the image could be rotated in any direction in the axial plane and the result would be similar). The four directions for analysis therefore include all possible directions from a reference pixel: horizontal, vertical, and both diagonals. By averaging across all directions we treat each direction of analysis equally and the result is spatially invariant. Through-plane directions were not used in this study due to the anisotropy of the voxels in our clinical images (5 mm slice thickness); such voxel anisotropy is typical of clinical CT follow-up imaging post radiotherapy.
<table>
<thead>
<tr>
<th>Equation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy $\sum_{i,j} g(i,j)^2$</td>
<td>Describes the uniformity of the image and is large for a constant image (all pixel intensity pairs are the same → high peak in the GLCM)</td>
</tr>
<tr>
<td>Entropy $-\sum_{i,j} g(i,j) \log_2 g(i,j)$ or 0 if $g(i,j) = 0$</td>
<td>Describes the randomness of the GLCM and is small for a constant image (all pixel intensity pairs are the same → high peak in the GLCM)</td>
</tr>
<tr>
<td>Correlation $\frac{\sum_{i,j} (i-\mu)(j-\mu)g(i,j)}{\sigma^2}$</td>
<td>Measures how correlated each pixel is to its neighbour and is high for a perfectly positively correlated image (the pixel intensity pairs are highly correlated in the GLCM)</td>
</tr>
<tr>
<td>Inverse difference moment $\sum_{i,j} \frac{1}{1 + (i-j)^2} g(i,j)$</td>
<td>Measures the contrast in the image and is high when there are large values on the GLCM diagonal (pixel pairs have equal intensities → peak along the GLCM diagonal)</td>
</tr>
<tr>
<td>Inertia $\sum_{i,j} (i-j)^2 g(i,j)$</td>
<td>Measures the contrast in the image and is low when there are large values on the GLCM diagonal (pixel pairs have equal intensities → peak along the GLCM diagonal)</td>
</tr>
<tr>
<td>Cluster shade $\sum_{i,j} ((i-\mu) + (j-\mu))^3 g(i,j)$</td>
<td>Describes the skewness (lack of symmetry) of the GLCM and is low when the image is symmetric with regard to its texture values (little variation in intensities)</td>
</tr>
<tr>
<td>Cluster prominence $\sum_{i,j} ((i-\mu) + (j-\mu))^4 g(i,j)$</td>
<td>Describes the skewness (lack of symmetry) of the GLCM and is low when the image is symmetric with regard to its texture values (little variation in intensities and peak in the GLCM around the mean intensity)</td>
</tr>
</tbody>
</table>
3.2.5 Data Analysis

Within the GGO regions, 9 features were computed: mean density, first-order texture (standard deviation [SD] of density), and the seven second-order texture features described previously (energy, entropy, correlation, IDM, inertia, cluster shade, and cluster prominence). Within the consolidation regions, the longest axial diameter and 3D volume were measured. Group differences were assessed using an independent samples t-test with unequal variances in SPSS Statistics Version 21.0 (IBM Corp., Armonk, NY, USA) to test the null hypothesis that the means of the recurrence and RILI groups are equal. A Kolmogorov-Smirnov test was also completed in SPSS to test for normality of distribution for all the measures. All statistical tests were two-sided with \( p \leq 0.05 \) indicative of statistical significance.

Figure 3-3 depicts the analysis performed on each of our features. To determine the predictive capabilities of each of these measures individually, classification was performed using the linear Bayes normal classifier [26-28] as implemented in the PRTools 4.2.1 (Delft Pattern Recognition Research, Delft, The Netherlands) [29] MATLAB toolbox. Each feature underwent 2-fold cross validation over 100 repetitions, with all cross validations stratified to each fold. The means and standard deviations of the classification errors, false negative rates (FNRs), and false positive rates (FPRs) were measured over the 100 repetitions. These were compared between feature sets using a Wilcoxon signed rank test for nonparametric data in SPSS. Receiver operating characteristic (ROC) curve analysis was also used to determine the predictive capabilities of each of these measures individually; specifically we calculated the area under the receiver operating characteristic curve (AUC) for each measure. Spearman rank
correlation coefficients were computed in SPSS to measure the correlations of all features in our data set.

**Figure 3-3:** Flow diagram showing analysis performed on each of the 11 features. This workflow was executed 11 times; once for each of the 11 features used in our study (2 size measures of the consolidation regions and 9 appearance measures of the GGO regions).

3.2.6 RECIST Progressive Disease Criteria for Prediction

For this study, the longest axial diameter was used as an imaging-based size measure for the consolidative regions. This measure forms the basis of the RECIST response criteria which are used as the current clinical standard for determining disease progression or response based on percentage change in the lesion diameter. To compare the accuracy of RECIST for predicting recurrence post-SABR with the performance of our imaging features, we classified as a recurrence each patient having a 20% increase in the longest axial diameter, and classified all other patients as RILI, according to the
current RECIST response guidelines. For each time period, the reference diameter was taken as the smallest previously measured diameter (either pre-treatment or post-treatment) [20]. For the 2–5 month time period, the reference diameter was the pre-treatment diameter. At the 5–8 month time period, the reference diameter was either the 2–5 month time diameter or the pre-treatment diameter, whichever was smaller.

3.2.7 Sensitivity of GGO Delineation

Although delineation of the consolidative regions is relatively straightforward for the operator (and computer-assisted delineation tools have been developed for this problem [30]), delineation of the GGO border requires more careful judgment and is substantially more time consuming. Thus, we sought to determine the sensitivity of classification performance to perturbations of the GGO boundaries. We performed a simulation study wherein manual GGO delineations were concentrically expanded and contracted in 1 mm increments, and classifier performance based on GGO features was re-evaluated for each such modification of the GGO delineations. Expansion of the GGO delineations was performed by varying a threshold on the distance transformation of the manually-delineated GGO region. Thresholds were set between 1 mm to 5 mm, at 1 mm increments. The expansion was performed such that the expanded GGO region did not intersect with the manually delineated consolidative region, chest wall, or mediastinum. Contraction was performed by thresholding the distance transformation of the binary complement of the manual delineation. Contraction was performed at 1 mm and 2 mm thresholds, as larger distances resulted in nearly complete volume loss of many small GGO regions.
3.3 Results

3.3.1 Group Differences

All samples passed a Kolmogorov-Smirnov normality test with $p > 0.05$. We report as follows all significant differences found; significant differences were not detected within either time period for all features not reported below. Within the 2–5 month follow-up time interval, the mean density within the GGO regions was significantly different between the RILI and recurrence groups ($p=.035$). Patients with recurrence had significantly denser GGO regions with a mean (±SD) density of $-602.01 ± 59.54$ HU compared to those with RILI of $-659.16 ± 55.93$ HU. The mean values of the second-order texture features including energy ($p=.036$), entropy ($p=.034$), and inertia ($p=.036$) were significantly different between the recurrence and RILI groups during this time period. Patients with recurrence had significantly larger entropy ($9.27 ± 0.65$) and inertia ($54.59 ± 25.85$) values, and significantly lower energy ($0.0021 ± 0.0015$) values compared to patients with RILI (entropy $= 8.66 ± 0.56$, inertia $= 33.42 ±15.84$, and energy $= 0.0034 ± 0.0012$); all $p < 0.05$.

Within the 5–8 month time interval, the mean first-order texture feature (measured as the SD of CT densities) within the GGO regions was significantly different between the RILI and recurrence groups ($p=.0019$). Patients with recurrence had a more variegated first-order texture within the GGO with a mean (± SD) variability of $213.63 ± 21.79$ HU compared to those with RILI of $173.52 ± 30.40$ HU. An illustrative example is shown in Figure 3-4, where a more variegated texture is demonstrated in (h) with a higher SD, compared to a smoother texture in (e) which has a lower SD. Patients with recurrence also had significantly denser consolidation regions ($p=.021$), mean density of -
84.66 ± 72.34 HU compared to -158.28 ± 65.42 HU for those with RILI. The mean values of the second-order texture features including energy (p=.031), entropy (p=.015), inertia (p=.018), and correlation (p=.0025) were significantly different between the groups at this time point. Patients with recurrence had significantly larger entropy (9.24 ± 0.55) and inertia (56.20 ± 17.24) values, and significantly lower energy (0.0020 ± 0.0012) and correlation (0.0069 ± 0.0012) values compared to those patients with RILI (entropy = 8.69 ± 0.34, inertia = 38.50 ± 13.80, energy = 0.0031 ± 0.00081, and correlation = 0.011 ± 0.0036).
Figure 3-4: Sample texture images and their corresponding second-order energy, entropy, and first-order (standard deviation [SD] of densities) texture values. Samples (a-d) show artificial examples and (e-h) show images taken from lung tissue. The plots show the change of each feature in the specified image relative to those in the leftmost image (a) or (e).
3.3.2 Classification Performance

Figure 3-5 shows the means and standard deviations of the overall errors, FNRs, and FPRs for each of the measures using 2-fold CV. The top three predictors were all second-order texture features within the GGO. GGO energy had the lowest prediction error of 23% (corresponding to 77% prediction accuracy) on 2-fold CV followed by GGO entropy and inertia with errors of 24-26%, compared to 40% and 42% error with the longest axial diameter and 3D volume respectively. The first-order measure also performed similarly with 27% 2-fold CV error, but had a larger standard deviation. The longest axial diameter and volume size measures were outperformed by all of the appearance measures, except for the cluster shade and cluster prominence second-order texture features. Of our top performing texture features, GGO energy had the largest FPRs but the lowest FNRs. GGO inertia had the most balanced FNR and FPR. Overall, the two size measures had the lowest FPRs, but highest FNRs.
Figure 3-5: (a) Cross validation errors, (b) false positive rates, and (c) false negative rates for 2-fold cross validation. Glyphs indicate mean values, the error bars indicating the standard deviation over the 100 repetitions.

In 2-fold cross validation, the errors, FPRs, and FNRs were compared for all possible feature pairings using a Wilcoxon signed rank test for nonparametric data in
SPSS (p ≤ 0.05 indicative of statistical significance) since not all of these samples passed a Kolmogorov-Smirnov normality test with p < 0.05. The size measures of longest axial diameter and volume had a significantly higher median 2-fold cross validation error when compared to each of the appearance measures, except each with cluster prominence. The longest axial diameter was also significantly different from the volume measure. There was no significant difference between the median errors for energy and entropy. First-order texture had a significantly different median error compared to all second-order features, except correlation and inertia. All other pairings were significantly different from one another. No significant difference was detected between the FPRs at 2-fold cross validation for energy vs. entropy, first-order texture vs. inertia, and first-order texture vs. longest axial diameter. There was also no significant difference detected between the FNRs for IDM vs. entropy, correlation vs. entropy, and cluster shade vs. cluster prominence. All other combinations were significantly different in terms of their median FPRs and FNRs.

The top two features in terms of 2-fold classification error were: GGO energy and GGO entropy. To explore these features in more detail and to compare them with size measures, scatter plots of the recurrence and RILI patients in feature space, as well as receiver operating characteristic (ROC) curves, are shown for GGO energy and entropy (Figure 3-6) and solid volume and longest axial diameter (Figure 3-7). Figure 3-6 indicate that the two classes are approximately equally separable in both feature spaces (by each feature alone), with few outliers from both classes, whereas Figure 3-7 indicates substantially poorer class separation by longest axial diameter and solid volume; the ROC curves shown in the figures corroborate these observations. The AUCs for all measures at
this 2–5 month time interval are shown in the first column of Table 3-3. GGO energy and correlation produced the largest AUCs of 0.81 (energy is shown in Figure 3-6b), whereas the size measures of longest axial diameter and volume produced AUCs of 0.716 and 0.652 respectively (shown in Figure 3-7c). GGO entropy and the first-order texture feature also performed similarly to energy and correlation in terms of AUC. Overall, energy was the best performer in terms of 2-fold cross-validation and AUC, with an error of 22.7% and AUC of 0.81.
**Figure 3-6:** (a) All lesions plotted by their energy and entropy values within the GGO at 2–5 months post treatment. Red circles indicate patients with recurrence and blue crosses indicate those with benign injury. Qualitative example images of the patients indicated by arrows are shown in Figure 3-9. (b,c) The corresponding ROC curves for the linear classifier on each feature alone; AUC = 0.809 for energy and AUC = 0.800 for entropy.
Figure 3-7: (a) All lesions plotted by their solid (consolidative volume) and longest axial diameter at 2–5 months post treatment. Red circles indicate patients with recurrence and blue crosses indicate those with benign injury. (b,c) The corresponding ROC curves for the linear classifier on each feature alone; AUC = 0.716 for longest axial diameter and AUC = 0.652 for 3D volume.
Table 3-3: Areas under the receiver operating characteristic curve (AUC) for each feature

<table>
<thead>
<tr>
<th>Feature</th>
<th>2 to 5 Months</th>
<th>5 to 8 Months</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy</td>
<td>0.809</td>
<td>0.875</td>
</tr>
<tr>
<td>Entropy</td>
<td>0.800</td>
<td>0.808</td>
</tr>
<tr>
<td>Inertia</td>
<td>0.782</td>
<td>0.800</td>
</tr>
<tr>
<td>First-order texture</td>
<td>0.791</td>
<td>0.867</td>
</tr>
<tr>
<td>Correlation</td>
<td>0.809</td>
<td>0.875</td>
</tr>
<tr>
<td>Density</td>
<td>0.764</td>
<td>0.667</td>
</tr>
<tr>
<td>Inverse difference moment</td>
<td>0.764</td>
<td>0.750</td>
</tr>
<tr>
<td>Cluster Prominence</td>
<td>0.709</td>
<td>0.758</td>
</tr>
<tr>
<td>Cluster Shade</td>
<td>0.618</td>
<td>0.675</td>
</tr>
<tr>
<td>Longest axial diameter</td>
<td>0.716</td>
<td>0.723</td>
</tr>
<tr>
<td>Volume</td>
<td>0.652</td>
<td>0.723</td>
</tr>
</tbody>
</table>

Due to the need to predict recurrence as early as possible after SABR treatment, we were most interested in features providing useful performance during the earlier 2–5 month time interval. To evaluate the repeatability of these measures at a later time interval on the same patients, we measured their performance during the 5–8 month time interval. The mean 2-fold cross validation error of GGO energy and entropy performed
within ±3% of the 2–5 month time interval. Longest axial diameter showed improvement in error measurements of 3%. In terms of the AUCs at the 5–8 month time period (shown in second column of Table 3-3), the top first and second-order texture features performed similarly or better compared to the 2–5 month interval. Longest axial diameter and volume also performed similarly to the 2–5 month interval.

3.3.3 Correlation of Measures

Spearman rank correlation coefficients were computed in SPSS to determine any dependence or relationship among or between the first and second-order texture features in our study. The correlations of GGO energy, entropy, and the first-order texture feature are shown in Table 3-4. We also measured the correlation of these texture features with the volume of the analyzed GGO region, as shown in Table 3-4. At 2–5 months post-SABR, the highest correlation was observed with energy and entropy, with a coefficient of -0.969; this correlation is apparent in Figure 3-6. In general, the second-order texture features were better correlated with each other than they were with the first-order measure. This pattern held at the 5–8 month time period, but overall all correlations decreased at this time point. We observed no statistically significant correlations of the texture features with the 3D volumes of the GGO regions.
Table 3-4: Spearman rank correlation coefficients and significance values for features within the ground-glass opacity

<table>
<thead>
<tr>
<th></th>
<th>2 to 5 Months</th>
<th>5 to 8 Months</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correlation</td>
<td>Significance</td>
</tr>
<tr>
<td>Energy vs. Entropy</td>
<td>-0.969*</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Energy vs. First-Order Texture</td>
<td>-0.936*</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Entropy vs. First-Order Texture</td>
<td>0.919*</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>GGO Volume vs. Energy</td>
<td>-0.171</td>
<td>0.457</td>
</tr>
<tr>
<td>GGO Volume vs. Entropy</td>
<td>0.062</td>
<td>0.788</td>
</tr>
<tr>
<td>GGO Volume vs. First-Order Texture</td>
<td>0.243</td>
<td>0.289</td>
</tr>
</tbody>
</table>

* Correlation is significant at the 0.01 level

3.3.4 RECIDIST Progressive Disease Criteria for Prediction

The overall accuracy of RECIDIST for predicting tumour recurrence at 2–5 months post-SABR was 52.2%, with a FNR of 25% and FPR of 72.7%. At 5–8 months post-SABR, RECIDIST provided an accuracy of 65.2%, with a FNR of 45.5% and FPR of 27.3%. To measure whether a recurrence threshold other than the 20% increase given by the current RECIDIST guidelines would provide better performance, we trained the linear
classifier on our entire data set, using the percentage increase in longest axial diameter. The classifier found optimal decision points higher than the RECIST guideline of 20%, placing the decision boundary at 49% for 2–5 months and 77% at 5–8 months with overall errors of 52.2% and 34.8%, respectively.

3.3.5 Sensitivity of GGO Delineation

Receiver operating characteristic curves were produced on the expanded and contracted GGO regions for classification using each of the top feature sets at the 2–5 month time interval. Figure 3-8 shows the sensitivity of the AUC for prediction of recurrence for our top first and second-order texture features based on these variations in GGO delineation. After contraction of the regions by 1 or 2 mm, the first and second-order texture features remained constant or decreased by 0.06 from their respective AUC values in the manual GGO. This decrease is likely due to the fact that many of the lesions had a small total volume of GGO, so a contraction of even 2 mm resulted in a large volume loss and the remaining region may not have provided sufficient voxel samples for a robust calculation of the texture features. With an expansion of the GGO up to 5 mm, the AUC values for all texture features remained constant with slight fluctuations across each level of expansion. First-order texture had fluctuations in its AUC up to 0.04, whereas the AUC for energy remained constant, throughout all levels of expansion. The AUCs for entropy were also stable, with a decrease of 0.06 at 3 mm, but returned to baseline at 4–5 mm. By 5 mm, all features remained within 0.04 of their respective baseline values from the manual GGO region. This demonstrates the robustness of the AUC measure at each level after GGO boundary contractions and expansions, suggesting robustness to variability in GGO delineation.
Figure 3-8: Sensitivity of the area under the receiver operating characteristic curve (AUC) for prediction of recurrence at 2-5 months post-SABR based on variations in GGO delineation. The vertical dashed line indicates the AUC values for the manual GGO delineation.

3.4 Discussion

Early detection of recurrence following SABR is critical to allow eligible patients to undergo salvage therapies, including surgery, chemotherapy, or possibly additional radiotherapy. Since CT scanning is the standard follow-up imaging modality, quantitative analyses of these images that are predictive of recurrence in individual patients at an early time point post-treatment could have great potential for translation to routine clinical use. This study has demonstrated the ability, based on texture measures of the GGO in CT images taken within 6 months of SABR treatment (within a 2–5 month interval following treatment), to predict patients’ eventual cancer recurrence after SABR treatment of early-
stage non-small cell lung cancer. This study also demonstrated that appearance changes in these regions precede changes in size of the consolidative/solid regions.

Previous studies have shown that current approaches to the interpretation of CT imaging are insufficiently accurate for clinical use in detection of local recurrence [31, 32]; accurate prediction of eventual recurrence within 6 months of treatment is supportive of salvage intervention. Although an increasing opacity on CT has shown to be indicative of recurrence, this feature is accurately predictive based on imaging acquired after 12 months post treatment [11]. FDG-PET alone or in combination with CT findings has also demonstrated the ability to detect recurrences more than a year post-treatment [33, 34]. However, one study has shown that residual FDG-PET uptake 12 weeks post treatment, with an SUV greater than or equal to 5.0, signifies an increased risk of local recurrence (sub hazard ratio for high SUV of 7.3) [35]. A recent study looked at texture features of the gross tumour volume on pre-treatment diagnostic CT scans (in a PET/CT context) of patients with stage IA–IV NSCLC being treated with SABR or conventional radiotherapy treatment [36]. Although the study did not find significant correlation of texture features with loco-regional recurrence, it was noted that motion artefact mitigation may be important to the prediction of local recurrence. In the context of the results of the Vaidya et al. study, our results suggest that assessment of the first follow-up scan, motion artefact mitigation, and analysis of the GGO region may be important to prediction of failure. To the best of our knowledge, this is the first study to show early prediction (within 6 months of SABR treatment) of eventual cancer recurrence based on measurements derived from CT imaging alone.
Our decision to use second-order statistical texture measures was motivated by a qualitative observation of differences between patient groups that appears to be related to vascularity in the GGO regions, ranging from smooth to variegated appearance through small numbers of larger vessels to larger numbers of small vessels. As these patterns may vary in terms of the number and magnitude of transitions from low to high density as one travels voxel-by-voxel through the image, we surmised that GLCM-based texture measures could capture such neighbourhood relationships explicitly, giving a relatively direct (and therefore understandable) quantification of this qualitative observation. In addition, there is previous literature showing the utility of GLCM-based texture features with lung pathologies [13, 14]. Thus, for this exploratory study, we limited our investigation to this small number of more established texture features, leaving a more comprehensive study of a larger number of texture features within the scope of future work in the context of a larger sample size.

We observed that second-order texture features within the GGO were the most accurate in predicting recurrence within the early 2–5 month time period. Measures of size, and appearance measures within the consolidative regions, were not as accurately predictive during the same time period. The most useful features in our study were second-order energy and entropy which were shown to be highly correlated. The utility of the first-order texture feature in differentiating recurrence from RILI also suggests the importance of overall intensity variations within the GGO along with local, second-order variations. The effectiveness of the second-order energy feature suggests that variations in the spatial uniformity of densities within the GGO may be important for prediction.
Energy and entropy are measures related to orderliness and how neighbouring pixels are organized. An ordered image is one in which the same pixel intensity pairs occur more frequently; this results in a few GLCM cells with higher values, compared to an unordered image where neighbouring pixels have different intensities throughout the image, resulting in more GLCM cells with lower values [37]. Energy increases with a more ordered image and, as indicated in Table 3-2, is weighted by the sum of squared elements in the GLCM. Entropy is weighted by the logarithm of the elements in the GLCM, which is also known as a statistical measure of randomness in the GLCM, decreases with a more ordered image, and is negatively related to energy. The difference between energy and entropy is in the weights put on the GLCMs, with energy increasing with orderliness and entropy increasing with disorder.

To illustrate the intuitive meanings of these features, Figure 3-4 shows sample texture images; artificial examples and qualitatively corresponding CT image samples of the lung. The resulting first-order texture, and second-order energy, and entropy values are shown for each example. A more ordered or constant image, for example reflecting a single vessel (Figure 3-4 a and e), results in relatively high energy and low entropy values, as compared to the other sample images in Figure 3-4. As the number of vessels increases, the first-order SD texture feature is either constant (Figure 3-4 a-c) or produces non-monotonic changes (Figure 3-4 e-g) but second-order energy is monotonically changing. In the lung image samples, as the number of vessels increases, energy consistently decreases and entropy consistently increases. However, note that in Figure 3-4 g and h, the different patterns in the lung are better differentiated with the first-order SD feature. Overall, we can observe that the first-order texture measure is sensitive to
overall changes in contrast, but is less sensitive to the spatial arrangement of contrasting elements (e.g. it does not differentiate well between one large vessel and multiple smaller ones, such as depicted in Figure 3-4 a-c). The second-order measures are sensitive to the spatial relationships of these contrasting elements, since they take pixel neighbourhoods into account.

We observed that patients with recurrence have lower values of energy and higher values of entropy, suggesting that GGO patterns closer to Figure 3-4 (c) and (g) within the GGO are predictive of recurrence. To illustrate the qualitative meaning of varying energy values in the context of the recurrences and RILI cases in our data set, Figure 3-9 shows four sample lesions from our study, sorted by increasing energy values in the GGO from (a) to (d). The lesion locations in feature space can also be observed in Figure 3-6, where Figure 3-9 (a) and (d) are at both extremes for energy values and thus easily separated, but Figure 3-9 (b) and (c) are much similar and harder to classify. The qualitative differences in GGO appearance between Figure 3-9 (a) and (d) are readily apparent, with the recurrence case in (a) having substantially greater variegation of image texture and contrast between the vasculature and the background, and (d) having an overall smoother GGO appearance. By comparison, the recurrence in Figure 3-9 (b) and the RILI in Figure 3-9 (c) are more difficult to qualitatively distinguish, and this is reflected in their more similar GGO energy texture values. We speculate that these observed differences (e.g. between Figure 3-9 (a) and (d)) may represent an increase in vasculature within these regions in the recurrence cases (e.g. Figure 3-9 (a)), perhaps indicating there are early vascular changes occurring which may eventually lead to a recurring tumour. Patients with RILI had a more homogenous GGO with fewer evident
vessels. At this early time point following treatment, this perhaps indicates the presence of fibrotic tissue and other cells involved in the pathological response of radiation pneumonitis following radiotherapy [38]. Future work including histological examinations of these regions would be useful toward elucidating the tissue composition of these regions and thus the meaning of our imaging observations.

**Figure 3-9**: The 2–5 month scans for four lesions in our study. The solid curves enclose consolidative regions and the dotted curves enclose ground-glass opacity (GGO) regions. Images of patients who had eventual cancer recurrence are shown in (a) and (b), and images of patients who developed RILI are shown in (c) and (d). The energy values in the GGO for these lesions progressively increased from (a) through (d). Their locations in feature space can be seen in Figure 3-6. (a) and (d) are at relative extremes in feature space on the recurrence and RILI sides, respectively, and are qualitatively distinct. (b) and (c) are examples of recurrence and RILI cases that are closer to the expected decision boundary, and are more difficult to distinguish qualitatively.

The results of this study must be considered within the context of its strengths and limitations. Although the sample size is small, this is one of the largest series of
recurrences in a study determining treatment outcomes post-SABR. Even with the small sample size, prediction accuracies remained robust using 2-fold cross validation; however, validation of all results on a larger data set must be performed and this study should be considered as hypothesis-generating. Intra and inter-observer variability in region of interest delineation was also not directly measured in this study; however, the robustness of our measures with expansion and contraction of the GGO suggests that variability in delineation has minimal impact on these measures for predicting recurrence. We also did not consider the effects of any co-morbidities, such as chronic obstructive pulmonary disease (COPD), respiration, or variations in blood flow on lung density [39]. Any variations between scanners and reconstruction techniques which may affect density measurements were also not considered [40]; however, all image acquisition parameters were set consistently to minimize any variations. In terms of the texture analysis, it may be valuable for future work to include a more extensive set of texture features and classifiers for analysis.

In conclusion, we have demonstrated the ability of first and second-order image texture features to predict eventual cancer recurrence based on CT images acquired within 5 months of SABR treatment with accuracies greater than 77% using a linear classifier. These appearance measures outperformed any measure of size during the same time interval, where accuracies of approximately 60% could be achieved. However, further validation of these results is needed on a larger sample size. We are also investigating a faster and more reproducible delineation or sampling method of the GGO regions to allow for clinical translation to a useful computer-aided diagnosis tool. The impact of these early measures on a clinician’s diagnosis of RILI versus recurrence, as
well as potential effects on clinical decision making must also be addressed in a user study. This could eventually prevent unnecessary, invasive, and risky procedures for patients with only benign disease and allow for the early salvage therapy for patients with recurrence.

3.5 References


Chapter 4

Towards the goal of understanding if semi-automated segmentation methods can be used in this decision support system, we evaluated a semi-automated graph cuts based approach to segment and sample regions of post-SABR radiological changes to be used to predict local recurrence following SABR.

The contents of this chapter were previously published in the Journal of Medical Imaging: SA Mattonen, S Tetar, DA Palma, AV Louie, S Senan, and AD Ward. Imaging texture analysis for automated prediction of lung cancer recurrence after stereotactic radiotherapy. J Med Imag 2015; 2(4):041010. Permission to reproduce this article was granted by the SPIE and is provided in Appendix A.5.

4 IMAGING TEXTURE ANALYSIS FOR AUTOMATED PREDICTION OF LUNG CANCER RECURRENCE AFTER STEREOTACTIC RADIOTHERAPY

4.1 Introduction

Stereotactic ablative radiotherapy (SABR) (also known as stereotactic body radiotherapy) is now the guideline-recommended treatment for patients with non-small cell lung cancer (NSCLC) who are medically inoperable or refuse surgery [1, 2]. Compared with conventional radiotherapy techniques, SABR involves the treatment of small lung tumours with higher doses per fraction in fewer fractions. Typically doses of up to 18 Gy per fraction are delivered in between 3-8 fractions over 1–2 weeks, in contrast to a dose of 2 Gy per day delivered over 4–6 weeks in conventional radiotherapy techniques. The higher doses used in SABR has led to local control rates of up to 90% at 3 years post-treatment, similar to those reported after surgery [3, 4]. However, radiation induced lung injury (RILI), such as radiation fibrosis, can occur after SABR. Some forms of benign changes can appear with similar size and shape to a recurring tumour on
computed tomography (CT) imaging, which is routinely acquired every three months as part of follow-up care. This confounds the critical clinical decision to provide potentially life-saving additional salvage therapies in cases where the cancer is recurring after SABR.

Present guidelines recommend the use of serial CT scans for follow-up after SABR, with the use of 18-fluorodeoxyglucose positron emission tomography (FDG-PET) imaging used only when appropriate [5]. FDG-PET is recommended when recurrence is suspected on CT; however, due to the high number of false positive findings on PET, patients eligible for salvage treatment should undergo a biopsy if feasible [5]. Qualitative image assessment has also been performed on CT images following SABR, and a high-risk feature set has been developed to discriminate benign fibrosis from recurrence [6, 7]. These features include an enlarging opacity, sequential enlargement from one scan to the next, a bulging margin, loss of linear margin, and air bronchogram loss. However, these features typically do not manifest until 1 year post-SABR [6].

Our overarching goal is to develop a fully automated system that will classify a CT image as recurrence or RILI, supporting the decision to prescribe salvage therapy to SABR patients with recurring tumours. This system will not require any manual delineation other than that which is performed during the normal clinical workflow, and will produce operator-independent, reproducible classification results. Quantitative radiomic image analysis has been increasingly utilized on CT, magnetic resonance imaging (MRI), and PET for differentiation of tumour types, grades, and for response assessment across many disease sites [8, 9]. Texture analysis has been investigated in predictive modeling of radiation pneumonitis after definitive lung radiotherapy [10]. It
has also been described for predicting the development of radiation pneumonitis after definitive radiotherapy for esophageal cancer [11, 12]. Cunliffe et al. compared radiologist-defined severity of normal tissue damage with CT texture features [11]. In an additional study, they demonstrated the ability to differentiate patients with and without clinical radiation pneumonitis by measurements of dose-dependent texture change between pre- and post-radiotherapy CT images [12]. To the best of our knowledge, these are the only papers present in the literature measuring CT texture for benign radiation induced lung injury. Our previous work presented in Chapters 3 evaluated quantitative CT image texture analysis for early prediction of recurrence after SABR [13, 14]. We have shown that second-order texture features based on grey-level co-occurrence matrices (GLCM) calculated within manually delineated ground-glass opacity (GGO) regions can predict recurrence within 6 months post-SABR. The regions of GGO refer to hazy regions of increased attenuation in the lung within which vascular regions can still be visualized, and these regions typically surround the consolidative mass. Texture features within these regions showed 2-fold cross-validation (CV) errors of 23–30% and areas under the receiver operating characteristic curve (AUC) of 0.78–0.81 [13]. As seen in Figure 4-1, patients with benign injury tended to have a smooth GGO appearance compared to a variegated appearance in patients with recurrence.

However, there are two main limitations to clinical translation of this technique. First, although inter-operator variability in manual segmentations on radiographic images is a well-known problem, little is known about predicting recurrence based on texture feature analysis within GGO segmentations performed by different operators [15]. As GGO boundaries are often barely discernible, it is reasonable to expect substantial inter-
operator variability. Second, manual 3D segmentation of the GGO is time-consuming, and automated GGO segmentation is extremely challenging due to the lack of any shape regularity and difficulty, even for the human medical expert, in judging the locations of the weak GGO boundaries. Inspired by our previous observation that a peri-consolidative region (defined by a concentric expansion of the consolidative mass) intended to sample GGO tissue yielded comparable classification performance to manually-delineated GGO [14], we conjectured that accurate classification performance could be obtained using an automatically-defined peri-consolidative region, rendering a complete GGO segmentation unnecessary.

Based on these observations and challenges, our primary objectives in the current study are as follows. First, we aim to measure the accuracy of texture features for predicting recurrence based on the first 3 month follow-up scan, with a peri-consolidative region derived from a semi-automatic segmentation of the consolidative region [16]. This decision support system would eliminate the need for any time consuming manual segmentations. Although the segmentation algorithm is semi-automatic, its only input is the Response Evaluation Criteria in Solid Tumours (RECIST 1.1) line segment, which is routinely obtained as part of the clinical imaging workflow; no additional user interaction is required [17]. We also aim to determine the reproducibility of the system’s segmentations recurrence predictions to inputs from different operators. Finally, we aim to compare the classification performance to manually delineated segmentations.
Figure 4-1: The 2–5 month follow-up image for four patients in our study. The solid lines encompass consolidative changes and the dashed lines encompass regions of ground-glass opacity, as delineated by a senior radiation oncology resident. The two lesions which developed benign radiation induced lung injury are shown in (A) and (B), and qualitatively the ground-glass opacity regions have a smooth appearance. (C) and (D) show two lesions that eventually developed cancer recurrence and in these images a variegated texture is visible in the ground-glass opacity regions.

4.2 Methods

4.2.1 Materials and Imaging

A total of 24 lesions from 22 patients treated for stage I NSCLC with SABR at the VU University Medical Center, Netherlands between February 2004 and February 2010 were used for this study. Of these 24 lesions, 11 were defined as local recurrences based on biopsy confirmation (8/11) and/or ultimate clinical outcome (3/11). The remaining 13 lesions developed moderate to severe radiological RILI CT changes based on expert assessment and had at least two years of imaging follow-up. These 13 RILI cases were chosen because they were especially challenging to distinguish from recurrences based on the first follow-up scan; for such cases computer-assisted decision support was deemed to have greatest potential value in the clinical context. The proportion of
recurrences is artificially inflated in this data set, and the true rate of local recurrence for a typical stage I NSCLC cohort is around 10% [3, 4]. During follow-up, 46 post-treatment diagnostic CT scans were taken at the first two follow-up time periods (2–5 and 5–8 months) post-SABR on one of three scanners at the VU Medical Center: Siemens Volume Zoom 4-slice, Siemens Sensations 64-slice (Siemens Nederland N.V., Den Haag, Netherlands) or Philips Brilliance iCT 256-slice (Royal Philips Electronics, Inc., Amsterdam, Netherlands). To eliminate the effect of image acquisition parameters on quantitative image analysis, all follow-up scans were acquired with the same acquisition parameters at inspiratory breath hold, 120 kVp, 100 mAs, spiral acquisition, 0.5 second rotation time, 2.5–5 mm slice thickness, with 70 cc of intravenous contrast (Ultravist-300; Bayer Pharma AG, Berlin, Germany) administered with a 25 second delay. Iterative reconstruction was not used in this imaging dataset and a standard sharp lung convolution kernel was applied (B60f) with 0.5–0.9 mm isotropic in-plane voxel dimensions.

4.2.2 Region of Interest Segmentation

4.2.2.1 Lung Regions

The regions of the image containing normal lung parenchyma can be obtained for the planning CT scan as part of the contouring normally performed during the radiation therapy planning workflow and mapped by image registration onto follow-up scans, or through the use of several automatic segmentation algorithms [18-21]. Although this step is outside of the scope of our method, for the purposes of reproducibility, we are providing the procedure that we followed to obtain this segmentation for each of our scans. Normal lung parenchyma was automatically segmented in ITK-SNAP (Version 2.4.0) using region competition snakes [22]. For trachea-adjacent tumors, the trachea was
separately segmented superior to the carina. A 3D rectilinear region of interest was defined encompassing the segmentation target (the entire lung volume or the trachea) and pre-processed using a sigmoid function, implemented as the intensity region filter in ITK-SNAP, with the threshold set above -300 HU and smoothness of 1.00. One or more spheres within the region of interest were used for initialization of the region competition snakes and were evolved using a sparse field level set [14, 23]. Parameters for the segmentation were set as follows: curvature velocity weight of 0.20, propagation force of 1.0, and 1500 and 250 iterations for the lung and trachea respectively. Post-processing was performed in MATLAB 8.4 (The Mathworks Inc., Natick, MA, USA) by a slice-wise morphological filling of small vessels or artifacts less than 10mm² within plane. This ensures larger vessels and any large consolidative components are not included in the normal lung parenchyma. A whole-lung volume was achieved by manually filling in the normal lung parenchyma to fill any abnormal consolidative regions.

4.2.2.2 Manual Segmentation of Consolidation and GGO

Segmentations were performed by a resident in radiation oncology (operator 1) on all follow-up CT images in ITK-SNAP 3.0 [24]. A lung window setting (level/window - 600/1500 HU) and the paintbrush tool or polygon outline were used. A mediastinal window setting (level/window 50/450 HU) was also used for tumours or fibrosis abutting the mediastinum. Consolidation changes were defined as an increased density with respect to normal surrounding lung with no visibility of vasculature within. GGO was defined as an increase in normal lung density with visibility of vasculature within. The segmentations were randomly checked by a thoracic radiation oncologist. The time
required to manually complete the consolidation and GGO segmentations was also recorded on a subset of images.

4.2.2.3 Semi-Automatic Segmentation of Consolidation

An overview of our semi-automatic segmentation approach can be seen in Figure 4-2. In the current clinical workflow, a physician measures the lesion’s longest axial diameter based on RECIST criteria to determine treatment response [17]. Since these measurements are taken during the normal clinical workflow by the physician; we use them as initializations for the segmentation algorithm. Specifically, the endpoints of the RECIST segment on the CT image serve as the only operator inputs to the segmentation algorithm. These line segments allow for localization of our post-SABR consolidative regions of interest on the follow-up scan eliminating the need for any deformable registration to localize the area from the planning scan.

The recently published OneCut graph cuts algorithm was used to segment the consolidative regions [25, 26]. This algorithm finds the segmentation that minimizes the L1 distance between unknown object and background appearance models. The variation of this algorithm that uses seeds for initialization minimizes the energy function

\[ E(S) = -\beta \| \theta^S - \theta^\bar{S} \|_{L1} + |\partial S|, \]  

(1)

where \( S \) is the segmentation, \( \theta^S \) and \( \theta^\bar{S} \) are the distributions of object and background intensities, respectively, and \( \beta \) is a tuning parameter (0.05 in our experiments) determining the relative contributions of the L1 intensity model difference and the segmentation perimeter length to the overall energy. In practice, \( \theta^S \) and \( \theta^\bar{S} \) are represented as histograms with a specified number of bins (64 in our experiments). The
tuning parameter was determined using a subset of 4 images (2 RILI and 2 recurrence) spanning differences in size, shape, and appearance of the consolidative regions. The input RECIST line segments used to tune the parameter was separate from the RECIST line segments used to validate the algorithm. The OneCut approach globally minimizes this energy function with a single graph cut and is particularly suited to our problem given its speed, its natural incorporation of object and background seeds as input, and its demonstrated superior performance to the closest competing GrabCut method [27] for bin counts > 20, allowing for segmentation of objects having more subtle intensity differences from background [25]. The foreground seeds were defined as voxels greater than or equal to a threshold of -200 HU within a sphere $A$, centered at the midpoint of the RECIST segment, with a diameter equal to the length of the RECIST segment plus 10 mm. The background seeds were defined as all voxels within normal lung parenchyma, not within sphere $A$, and within a sphere $B$ centered at the midpoint of the RECIST segment with a diameter equal to the length of the RECIST segment plus 20 mm. The OneCut segmentation was performed on each slice (mapping the lung window/level range of -1350 to 150 HU to an 8-bit range) within a region of interest (ROI) centered on and enveloping the background seeds. Any parts of these segmentations lying outside of the whole lungs (in cases where the consolidation abutted the lung boundary) were removed and the 3D largest connected component closest to the RECIST line was taken. Due to the inclusion of small adjacent vessels in the 3D volume, a subsequent slice-by-slice 2D largest connected component was taken to remove these disconnected vessels in plane. Finally, the 3D largest connected component was taken as the final consolidation segmentation volume.
Figure 4-2: An overview of our semi-automatic segmentation approach and the methods used in this chapter. On the original CT image, an operator places a RECIST line segment to measure the longest axial diameter (taken in the normal clinical workflow). This was used as initialization for the OneCut algorithm to automatically obtain the consolidation segmentation (red). An expansion of this region defined the peri-consolidative region (blue). On the original CT image manually delineated consolidation and ground-glass opacity (GGO) regions were also obtained. Texture analysis was performed in the GGO and semi-automatic peri-consolidative regions, and classification results were compared using 2-fold cross validation errors, false positive rates (FPR), false negative rates (FNR), and area under the receiver operating characteristic curve (AUC).

To validate our algorithm and assess the impact of inter-operator variability on segmentation and classification performance, three operators provided RECIST measurements of the consolidative changes post-SABR. If the consolidation region was split into more than one disconnected region, a separate measurement was taken for each region. Validation was performed with typical users of the system, including a senior radiation oncology resident (operator 1) and two thoracic radiation oncologists (operators
2 and 3). To avoid biasing the results, none of the operators in this study contributed to the design of the semi-automated segmentation algorithm.

4.2.2.4 Automatic Delineation of Peri-Consolidative Region

A peri-consolidative region within the lung parenchyma was derived automatically by thresholding the 3D distance transform of the consolidation segmentations. The threshold was set at 16 mm based on our previous observation that classification performance does not differ substantially from that given by the corresponding manually delineated GGO above this threshold [14]. The domain was also restricted to a sphere, centered at the midpoint of the RECIST segment, with a diameter equal to the length of the RECIST segment plus 32 mm (16 mm × 2). This is due to the possible inclusion of connected vessels distant to the consolidative mass or extra consolidative regions erroneously included in the semi-automatic segmentation. A sphere defined by the RECIST line segment will, by definition, circumscribe the lesion; this shape was chosen to enable concentric sampling of the tissue outside of the consolidative mass. We want to ensure we are sampling the peri-consolidative region adjacent to the treatment site and avoiding sampling additional normal lung distant to the consolidative mass. The size of the sphere was chosen to encompass the entire RECIST line plus a 16 mm margin on each end, which was chosen for consistency with the 16 mm threshold used for expansion from the consolidative mass. An example of a resulting peri-consolidative region is seen in Figure 4-2.

4.2.3 Feature Extraction and Image Analysis

MATLAB 8.4 (The Mathworks, Natick, MA) was used to calculate first-order texture as the standard deviation of the density within the GGO or peri-consolidative
region. The Insight Segmentation and Registration Toolkit (ITK) 4.3.1 (www.itk.org) was used to calculate four second-order texture features based on a gray level co-
ocurrence matrix (GLCM): energy, entropy, inertia and correlation [28-30]. The equations for calculation of these texture features are

\[ \text{Energy} = \sum_{i,j} g(i,j)^2, \]  
\[ \text{Entropy} = -\sum_{i,j} g(i,j) \log_2 g(i,j) \text{ or } 0 \text{ if } g(i,j) = 0, \]  
\[ \text{Correlation} = \sum_{i,j} \frac{(i-\mu)(j-\mu)g(i,j)}{\sigma^2}, \]  
\[ \text{Inertia} = \sum_{i,j} (i-j)^2 g(i,j), \]

where \( g \) is a two-dimensional matrix where each element \( g(i,j) \) contains the number of voxel pairs whose elements have gray levels \( i \) and \( j \), where \( \mu \) is the weighted pixel average

\[ \mu = \sum_{i,j} i \cdot g(i,j) = \sum_{i,j} j \cdot g(i,j) \text{ (due to symmetry of } g), \]

and \( \sigma \) is the weighted pixel variance

\[ \sigma = \sum_{i,j} (i-\mu)^2 \cdot g(i,j) = \sum_{i,j} (j-\mu)^2 \cdot g(i,j) \text{ (due to symmetry of } g). \]

The number of bins and density ranges in the GLCM were set to yield 20 HU bin widths between -1000 HU and 200 HU (60 bins), based on analysis of the histogram distributions within manually delineated GGO regions. GLCMs were calculated within four in-plane neighbouring voxel pair directions \[\{-1, 0, 0\}, \{-1, -1, 0\}, \{0, -1, 0\}, \{1, -1, 0\}\] for the entire 3D region of interest and texture features were averaged over all directions.

When calculating a GLCM, the neighbouring voxel to be analyzed must be specified by
the distance and location (in the x, y, and z directions) from the reference voxel. An example of the four in-plane neighbouring voxel relationships is shown in Figure 4-3. Due to the voxel anisotropy typically seen in post-radiation follow-up lung CT images (5 mm slice thickness), we did not analyze through-plane directions in this study. Features were analyzed within two discrete time periods: 2–5 and 5–8 months post-SABR. This timing of images was chosen as the focus of our study is on the early prediction of recurrence post-SABR. Within each time period, the images used for analysis spanned all 22 patients and 24 tumours available.

\[
\begin{array}{ccc}
(-1, -1, 0) & (0, -1, 0) & (1, -1, 0) \\
(-1, 0, 0) & & \\
\end{array}
\]

**Figure 4-3:** The four in-plane spatial relationships used for calculating the gray level co-occurrence matrix (GLCM). The dark central voxel is the reference voxel, and the neighbouring voxels considered in the analysis are shown in gray.

### 4.2.4 Classification

PRTools 5.0 (Delft Pattern Recognition Research, Delft, The Netherlands) was used for classification. For a stringent determination of classification performance, two-fold CV over 1,000 repetitions was performed using a linear Bayes normal classifier. The mean and standard deviation of the classification error, false negative rates (FNRs), false positive rates (FPRs), and AUCs were measured, where recurrence is defined as positive. Classification performance was measured for all five extracted texture features in both
the automatically-defined peri-consolidative regions and the manually-delineated GGO regions.

4.2.5 Comparison of the Semi-Automatic Segmentations

To compare the segmented regions between operators, similarity metrics were calculated to measure segmentation differences. The symmetric mean absolute boundary difference (MAD), Dice similarity coefficient (DSC) \( \frac{2 \cdot |V_A \cap V_B|}{|V_A| + |V_B|} \), volume difference (VD) \( V_B - V_A \), recall \( \frac{TP}{TP + FN} \), and precision \( \frac{TP}{TP + FP} \) were calculated, where \( V_A \) and \( V_B \) are the volumes of the segmentations by operator A and B, respectively, and \( TP = \text{true positive}, \ FP = \text{false positive}, \ TN = \text{true negative}, \) and \( FN = \text{false negative} \). Segmentation metrics were calculated on all of the images used in this study. Each metric was calculated on the same 3D image volume between operators. Calculation of segmentations differences was completed between all operators for the semi-automated consolidative regions. Due to the lack of a reference operator when comparing segmentations between operators, the \( F_1 \) score was calculated to eliminate the effect of reference operator selection on precision and recall metrics. The \( F_1 \) score is the harmonic mean of the precision and recall and can be used to measure the segmentation accuracy \( \frac{2 \cdot (\text{precision} \cdot \text{recall})}{\text{precision} + \text{recall}} \). Additionally, the time to generate the manual and semi-automatic segmentations was measured and compared.

4.2.6 Statistical Analysis

All statistical analyses were completed in SPSS Statistics Version 22.0 (IBM Corp., Armonk, NY, USA). To compare segmentation differences between operators, a
Kolmogorov-Smirnov test was first completed to test for normality of distribution for all the measures. A Wilcoxon signed rank test for nonparametric data was used to compare differences between operators (two-sided with alpha ≤ 0.05). To compare classification performance between operators, an independent samples t-test with unequal variances was performed to test the null hypothesis that the mean classification performance between operators was equal. Due to the repeated sampling in the cross-validation metrics (1000 times), to correct for multiple testing a Bonferroni correction was applied with alpha ≤ 0.05/1000. To determine non-inferiority of the semi-automated classification results with respect to results obtained using the manual segmentations, the 95% confidence interval of the difference in cross-validation metrics was assessed. An inferiority margin of 5% was chosen as an acceptable clinical difference in cross-validation metrics.

4.3 Results

4.3.1 Manual Segmentations

Classification performance at 2–5 months post-SABR using texture features within manually delineated GGO regions is shown in Figure 4-4. We examined classification results within this early time range, as we want to determine the ability to predict recurrence as early as possible post-SABR. The top-performing feature in terms of 2-fold CV error was entropy. At 2–5 months post-SABR, the mean 2-fold CV error was 27.4%, with a mean FPR and FNR of 29.5% and 25.3% respectively, and an AUC of 0.64. For comparison, the AUC value at 5–8 months post-SABR was 0.67 and is shown in Table 4-1. A 5–8 months post-SABR, the entropy feature demonstrated a 2-fold CV
error of 30.0%, mean FPR of 25.2% and mean FNR of 36.2%. A qualitative representation of the manually delineated regions of interest is provided in Figure 4-1.

**Figure 4-4:** Classification performance of the manual GGO and semi-automatic per-consolidative regions for the top-performing texture feature (entropy) at 2–5 months post-SABR. The columns indicate the mean 2-fold cross-validation errors, mean false positive rates (FPR), mean false negative rates (FNR), and the area under the receiver operating characteristic curve (AUC). The whiskers indicated the standard deviation over 1,000 repetitions and each color represents a different operator. The asterisks indicate a statistically significant difference between operators (p < 0.00005).
Table 4-1: Area under the receiver operating characteristic curve for the top texture feature (entropy) at both time points post-SABR

<table>
<thead>
<tr>
<th>Operator</th>
<th>2–5 months</th>
<th>5–8 months</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operator 1 (Manual)</td>
<td>0.64 ± 0.04</td>
<td>0.67 ± 0.04</td>
</tr>
<tr>
<td>Operator 1 (Semi-auto)</td>
<td>0.73 ± 0.06</td>
<td>0.71 ± 0.05</td>
</tr>
<tr>
<td>Operator 2 (Semi-auto)</td>
<td>0.70 ± 0.04</td>
<td>0.67 ± 0.07</td>
</tr>
<tr>
<td>Operator 3 (Semi-auto)</td>
<td>0.71 ± 0.05</td>
<td>0.69 ± 0.07</td>
</tr>
</tbody>
</table>

4.3.2 Semi-Automated Segmentations

4.3.2.1 Segmentation Comparison

Similarity metrics for the semi-automatic consolidative segmentations are summarized in Table 4-2. The majority of the metrics failed a Kolmogorov-Smirnov test for normality, and therefore the median values are reported. A Wilcoxon signed rank test for nonparametric data was performed to test the null hypothesis that the medians of each group are equal. Overall, when using the semi-automatic approach, inter-operator variability was low with values for the F1 score ranging from 0.87–0.93 and boundary differences from 1.00–1.75 mm. Operator 3 showed the largest variability with significant differences in all four metrics. However, the volume overlap (DSC) measurements still showed high overall agreement in the segmentations, with values of 0.87–0.93. Qualitative examples for all of the operators’ semi-automatic segmentations are shown in Figure 4-5.
**Table 4-2**: Inter-operator variability in the semi-automatic consolidative segmentations; all values are reported as the median [interquartile range].

<table>
<thead>
<tr>
<th>Operator</th>
<th>MAD bilateral (mm)</th>
<th>DSC</th>
<th>F1 Score</th>
<th>Volume Difference (cm³)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 vs 2</td>
<td>1.18 [0.31, 2.62]</td>
<td>0.90 [0.78, 0.98]</td>
<td>0.90 [0.78, 0.98]</td>
<td>-1.66# [-17.91, 0.33]</td>
</tr>
<tr>
<td>1 vs 3</td>
<td>1.75* [0.60, 3.19]</td>
<td>0.87* [0.74, 0.97]</td>
<td>0.87* [0.74, 0.97]</td>
<td>-10.56*# [-22.86, -1.68]</td>
</tr>
<tr>
<td>2 vs 3</td>
<td>1.00* [0.45, 2.13]</td>
<td>0.93* [0.81, 0.97]</td>
<td>0.93* [0.81, 0.97]</td>
<td>-3.57* [-13.58, 0.06]</td>
</tr>
</tbody>
</table>

*#Indicates significant differences between rows, p < 0.05

**Figure 4-5**: The pre-treatment lesion used for treatment planning and a 3-month follow-up scan showing the radiological changes post-SABR for a patient with radiation-induced lung injury (1) and recurrence (2). Semi-automated segmentations of the consolidative regions (red) and peri-consolidative regions (blue) surrounding them. Operators 1, 2 and 3 are shown for each case.
4.3.2.2 Classification Performance

Classification performance at 2–5 months post-SABR within the automatically derived peri-consolidative regions is shown in Figure 4-4 for the top feature (entropy). Using these segmentations, entropy had mean 2-fold CV errors of 24.7-27.8% across all three operators. Operator 1 had a significantly lower mean 2-fold error compared to operators 2 and 3. There was no significant difference in mean errors between operators 2 and 3. Overall, the decision support system based on semi-automatic segmentations produced balanced FPRs and FNRs of 25.8-30.6% and 23.5-25.3% respectively. There was no significant difference in the mean FNRs between any of the operators. All three operators had significantly different mean FPRs, however differences were all ≤ 5%. At 2–5 months post-SABR entropy had AUC values between 0.70-0.73 for all operators, which was significantly different between all operators. For comparison, the AUC values at 5–8 months post-SABR are shown in Table 4-1 and were between 0.67–0.71 for all operators. At this time point, the entropy feature had 2-fold CV errors of 22.8-31.3%, mean FPRs of 21.6-23.3% and mean FNRs of 22.8-43.9%.

4.3.3 Comparison of System Performance: Manual vs Semi-Automatic Segmentations

4.3.3.1 Classification Performance

Comparison of classification performance at 2–5 months post-SABR using the manual and semi-automatic decision support systems is shown in Figure 4-4. Classification using the semi-automated method was compared to classification performance using the manual delineations performed by the senior radiation oncology resident. Using the semi-automatic segmentations, the entropy texture feature produced
AUCs higher than the manual contours for all three operators. All AUC values using the semi-automated approach were significantly different from reference manual contours. In terms of mean 2-fold error, operator 1 demonstrated a significantly better performance compared to their manual segmentations. There was no significant difference in classification performance between the manual and the semi-automatic segmentations by operators 2 and 3. There were also no significant differences with respect to the FNRs. However, operators 1 and 3 produced significantly lower FPRs compared to the manual segmentations, whereas no difference was observed with respect to operator 2. A non-inferiority study was completed to determine if the semi-automated approach was non-inferior to the manual approach. Figure 4-6 shows the 95% confidence intervals of the differences between the semi-automated and manual classification metrics. Non-inferiority was demonstrated for all metrics as the 95% confidence intervals fell within clinically acceptable differences of 5%. At 5–8 months post-SABR the AUC and 2-fold errors values were similar to 2–5 months. Results were also consistent between the manual and semi-automated systems, with AUC values within 0.04 and 2-fold errors within 7%. Overall, classification results showed robustness using the semi-automatic approach and were comparable to or exceeded the results using manually delineated regions. Individual feature values for the recurrence and injury groups, as well as all classification metrics are summarized in Appendix C, Table C1 and C2.
Figure 4-6: Non-inferiority analysis for the difference between means (semi-automated – manual) for the classification error metrics (A) and area under the receiver operating characteristic curve (AUC) (B). The markers indicate the mean differences and the whiskers represent the 95% confidence intervals. The vertical dashed lines represent an acceptable clinical difference of 5% for all metrics.
4.3.3.2 Timing

The average time (± SD) for operator 1 to manually delineate the consolidative and GGO regions on each image (for a subset of 20 images) was 579 ± 472 seconds (i.e. 9.6 ± 7.9 minutes). The semi-automatic approach took 27 ± 25 seconds to obtain the consolidative and peri-consolidative regions on each image. The increase in segmentation speed was 20-fold, with an average savings of 9 minutes per image.

4.4 Discussion

The ability to distinguish benign fibrosis from tumor recurrence is crucial in determining a patient’s care following SABR, and determine whether salvage surgery or additional radiotherapy is required. Current clinical guidelines recommend the use of serial CT imaging for follow-up assessment after treatment with SABR. Therefore, a reliable measure for determining recurrence on CT imaging would be extremely valuable as the utilization of SABR is rapidly increasing. The use of quantitative appearance measures could provide an early assessment of response through quantifying subtle patterns predictive of recurrence not typically considered by a radiologist or radiation oncologist.

Our work in Chapter 3 has shown that the entropy texture feature calculated within manually delineated regions of GGO could predict recurrence with 2-fold CV errors of 24% at 2–5 months post-SABR [13]. The results presented in this study using manual segmentations by a different operator were concordant, with an error of 27%. There exist radiographic changes to the tissue surrounding the consolidative regions as a result of SABR (as shown in Figure 4-1), which can cause a substantial loss of boundary contrast. These observations provide important context for the evaluation of automatic
segmentation algorithms for consolidation regions on post-SABR lung CT images by quantifying the uncertainty inherent in the manual reference standard. GGO regions can also have a highly variable appearance and an ill-defined border, rendering them very challenging to delineate. This emphasizes the difficulties in delineating these regions and lack of a single ground truth reference standard segmentation for this problem. To eliminate the need for time consuming manual segmentations and any inherent variability between them, the goal of this study was to produce an accurate and reproducible means of recurrence prediction post-SABR using semi-automatic segmentations.

This work has shown the ability to predict recurrence post-SABR using texture analysis in regions delineated by means of a semi-automated segmentation algorithm, initialized using only the RECIST diameter measurement that is normally collected during the clinical workflow. This diameter measure on post-SABR follow-up scans is assessed according to the RECIST 1.1 criteria by comparing it to the pre-treatment scan [17]. Consequently, it provides a quick and efficient means for initialization of our decision support system. It will also guarantee localization of our post-SABR consolidative regions of interest regardless of deformations and retractions from the pre-treatment location. We also considered that in the typical clinical workflow, the RECIST line segment would be taken by the radiologist reading the CT image and would be external to our workflow. In other words, we would obtain this line segment when we obtain the follow-up image, leading to a fully automatic overall system based on semi-automated segmentation.

We reported small differences between the semi-automated segmentations obtained from each operator. This is due to variability in the placement of the RECIST
line segment. It has been well described that substantial variability exists, both between and within operators, in the measurement of response in lung cancer patients [31, 32]. This variation can be seen in the placement of line segments measuring longest axial diameter according to RECIST guidelines. This can be particularly evident in the post-SABR context due to the highly variable size and shape of the consolidative masses (they are not typically spherical as is commonly the case for pre-treatment tumors) as shown in Figure 4-1. Consequently, it can be difficult to determine the correct location on which to place the line segment. However, the differences we observed in the physical placement of the line segment had a minor impact on the graph cuts algorithm, since the foreground and background samples (as shown in Figure 4-2) are likely to vary only slightly. The $F_1$ scores were all greater than 0.85, suggesting high inter-operator agreement in the semi-automatic consolidative regions.

As mentioned previously, the GGO surrounding the consolidative regions on follow-up images decreases the boundary contrast of the consolidation. Overcoming the loss of contrast using smoothness and/or shape priors may not be straightforward, given the non-smooth and highly variable shapes of the consolidative regions (as seen in Figure 4-1). Most importantly however, this semi-automatic segmentation provides basis for the automatically-defined peri-consolidative region. As shown in Figure 4-5, the peri-consolidative regions are in general dissimilar to the manual GGO regions seen in Figure 4-1. This is intentional, as our intention is to sample a region surrounding the consolidative regions within which to calculate the texture features. Our previous work demonstrated that sampling a region surrounding the consolidative region was sufficient
for predicting recurrence, and that a complete segmentation of GGO was not required [14].

Our system shows high predictive accuracy using the second-order entropy texture feature with AUCs of 0.70–0.73 at 2–5 months post-SABR. It was also robust to inter-operator variability in the initialization of the system by placement of the RECIST diameter measurement. Although classification performance using the semi-automated method was compared only against a single operator’s manual segmentations, non-inferiority was demonstrated among all three operators. Robustness in classification performance was also demonstrated among operators, suggesting a reproducible means for delineating the consolidative regions even with variations in RECIST line segment placement among operators. Our semi-automatic approach potentially eliminates the need for manual segmentations and this provides an important avenue for future work on a larger dataset as part of our ongoing work. Results at 5–8 months post-SABR were comparable to the results at 2–5 months post-SABR. This demonstrates stability of system performance through time, and also demonstrates that the appearance of radiological changes at the two earliest clinically scheduled follow up time points (approximately 3 and 6 months) could be important in predicting recurrence. As previously demonstrated in patients with recurrence, these regions have a more variegated texture, perhaps indicative of early vascular changes as seen in Figure 4-1. Patients who develop only benign fibrosis seem to have a more homogenous appearance to the GGO, or minimal appearance of GGO changes post-SABR. Our ongoing work is examining these regions of post-SABR changes histologically, to determine their composition and correlate this to observations on CT imaging.
Previous studies using qualitative high-risk appearance features have shown the utility of categorizing a patient's risk of recurrence, however most of these features do not typically appear until a year post-treatment [6]. Other imaging-based features such as size or mass-like shape can be ineffective for early detection of local recurrence post-SABR [33-35]. FDG-PET has also been investigated for distinguishing fibrosis from recurrence post-SABR and it has been shown that maximum standardized uptake values (SUV\textsubscript{max}) can be predictive of recurrence, but not until a year post-SABR [36, 37]. A recent approach using CT perfusion imaging for response assessment has been investigated for pulmonary metastases undergoing SABR; however, further analysis is needed in a larger cohort of patients [38]. The combination or addition of these techniques to our decision support system may be useful for aiding in the assessment of difficult cases.

To the best of our knowledge, there has been no previous study on semi-automatic segmentation of post-SABR consolidative changes on CT. An interesting avenue for future work would be in refining these segmentations to enable quantification of shape and size changes of these regions. Another interesting area of further study could look at additional methods for segmentation. One such possibility is using the tumour delineation on the planning scan and performing a deformable registration to the follow-up scan. Such an approach would be fully automated but would require highly accurate deformable registration capable of compensating for highly variable and localized post-SABR radiological changes. This would also need to compensate for tissue retractions that can occur in the lung post-SABR, displacing the consolidations from their original locations in the planning scan. Nevertheless this would be an interesting future study and would be considered a fully automated method, eliminating the impact of variability in
RECIST line segment placement by different physicians. However, our initial work has shown robustness of the decision support system to differences in line segment placement by different operators.

We must consider our study in the context of its limitations, including the small sample size of patients with significant benign fibrosis. Our ongoing work involves validation on a larger sample set of patients matched based on patient and treatment characteristics. This will allow for a more comprehensive analysis of post-SABR changes. This study also focused solely on our previously published top five texture features for recurrence prediction, and further work on a more compressive radiomic feature set and machine learning platform may improve prediction results on a larger data set. The focus of this study was on the early prediction of recurrence and images analyzed at 3 and 6 months post-SABR. Further studies should be completed on additional time points to determine the usefulness of image features for recurrence prediction as time post-SABR increases and to determine the optimal time point for prediction. This study also did not consider the effect of different scanners and reconstruction techniques in the analysis; however, all images were taken with the same acquisition and reconstruction parameters to minimize discrepancies [39]. Determining the effect of different acquisition parameters on classification results is an interesting avenue for future study. Also, reference manual contours for classification were completed by a single operator and the validation of our semi-automated algorithm was completed with only 3 operators with similar expertise. Further validation should be completed with additional operators with different expertise (ex. radiologists or senior radiation oncologists) who would be considered typical users of this time of algorithm. Another limitation of the current work
is the use of the same imaging dataset as used in Chapters 2 and 3. Using the same small dataset could potentially introduce bias into our results. To enable clinical translation, a comprehensive validation and user study on a larger dataset is ongoing.

4.5 Conclusion

Second-order texture features calculated within GGO delineated from a semi-automated algorithm, initialized using only the RECIST diameter measurement routinely taken during the clinical workflow, have shown the potential to predict recurrence in individual patients within 6 months of SABR. At 2–5 months post-SABR, second-order entropy provided good recurrence prediction based on semi-automatic segmentations, with AUCs of 0.70–0.73; the corresponding result using a manual segmentation was 0.64. This system demonstrated consistent segmentations and prediction accuracies between operators, which were concordant with prediction accuracies, based on a single reference manual segmentations and obtained 20 times faster using the automated approach. The next step of this study is to validate our algorithm on an additional 93 patients we have obtained. This work has the potential to lead to a clinically useful computer-aided diagnosis tool which can be easily integrated into a physician’s workstation and eliminate the need for any manual segmentation. An automated decision support system can improve the physician’s assessment of response following SABR to predict recurrences as early as possible. This will allow patients to receive timely salvage therapies, and reduce the risk of patients with only benign fibrosis undergoing risky biopsy procedures.

4.6 References


Chapter 5

Towards the goal of understanding how physicians perform in response assessment post-SABR, we evaluated and compared a radiomics decision support system to physician assessment of response following SABR.


5 DETECTION OF LOCAL CANCER RECURRENCE AFTER SABR FOR LUNG CANCER: PHYSICIAN PERFORMANCE VERSUS RADIOMIC ASSESSMENT

5.1 Introduction

Stereotactic ablative radiotherapy (SABR) is a guideline-recommended treatment option for patients with early-stage lung cancer who are medically inoperable or refuse surgery [1, 2]. SABR can achieve local control rates comparable to surgery of 90% at 3 years post-treatment [3]. However, following treatment with SABR, patients often develop benign radiation-induced lung injury which can mimic recurrence. In some cases this leads to resection for lesions ultimately proven to contain only fibrotic tissue [4-7]. Early detection of patients with local recurrence using standard follow-up computed tomography (CT) images may allow for timely salvage surgery, which has been shown to be feasible in such patients [8, 9]. Accurate recurrence detection may avoid unnecessary scans and interventions in patients harboring only benign fibrosis.
Local recurrences typically manifest at a median time of 15 months following SABR, but may present up to 5 years later [10]. Qualitative high-risk CT features (HRFs) have been validated as predictors of recurrence, and include the presence of an enlarging opacity, enlargement after one year, sequential enlargement from one scan to the next, bulging margin, linear margin disappearance, air bronchogram loss, and cranio-caudal growth [11-13]. However, such HRFs are subject to inter-observer variability, and may not be observable until beyond 1-year post treatment [12].

Radiomics is a developing area of research that aims to extract more complex information from conventional medical images, including features not easily visible or quantifiable with the naked eye. These features can be used to build models for clinical outcomes, including diagnostic, prognostic, or predictive information [14, 15]. We hypothesize that radiomic image features could improve prediction accuracies by detecting subtle imaging changes not observed by a physician. In recent pilot studies, we demonstrated the ability of quantitative CT image analysis for early prediction of recurrence after SABR [16-18]. Texture features calculated within areas intended to subsample post-SABR ground-glass opacity could predict recurrence at 3-months post-SABR, with areas under the receiver operating characteristic curve (AUC) of 0.78–0.81. However, a limitation to clinical translation of this technique was the small dataset and limited set of radiomic features used.

To our knowledge, no prior studies have assessed either physician performance or inter-physician variability and reliability in detecting recurrence after SABR. The goals of this study were to determine the accuracy of expert radiologists and radiation
oncologists’ assessment of local recurrence versus non-recurrence on a series of follow-up CT images post-SABR, and to compare physician performance to a radiomics tool.

5.2 Methods and Materials

5.2.1 Materials and Imaging

The study population consists of patients with T1/T2N0 non-small cell lung cancer (NSCLC) treated with SABR. 15 patients with local recurrence were matched 1:2 with patients without recurrence, but with radiological radiation-induced lung injury (RILI), according to baseline factors including planning target volume (PTV) size, tumour location, and fractionation as previously described [12]. These tumours were treated to a dose of 54–60 Gy in 3–8 fractions, using a risk-adapted approach, as described previously [19]. Of the 15 patients with local recurrence, seven recurrences were confirmed with biopsy and the remaining eight were determined local recurrences based on a combination of CT, FDG positron emission tomography (PET) findings and subsequent clinical outcomes. The maximum inspiration phase of respiration was used for analysis on the pre-treatment planning CT image. Standard follow-up practices were to acquire diagnostic CT images at 3, 6, and 12 months post-SABR and every 6-12 months thereafter, although the actual timing of scans were subject to normal clinical variability. Matched patients were required to have follow-up CT imaging at similar time intervals and durations.

Pre-treatment CT scans used for treatment planning were acquired for each patient using four-dimensional (4D) CT (Lightspeed 16; GE Medical Systems, Waukesha, USA) at 140 kVp, 100–110 mAs, and 2.5 mm slice thickness. The maximum
inspiration phase of respiration was used for analysis. Post-treatment diagnostic CT scans were taken on one of three scanners: Siemens Volume Zoom 4-slice, Siemens Sensations 64-slice (Siemens Nederland N.V., Den Haag, Netherlands) or Philips Brilliance iCT 256-slice (Royal Philips Electronics, Inc., Amsterdam, Netherlands). Follow-up scans were acquired at inspiratory breath hold with standard imaging parameters of 120 kVp, 100 mAs, spiral acquisition, 0.5 second rotation time, 2.5–5 mm slice thickness, and 70 cc of intravenous contrast (Ultravist-300; Bayer Pharma AG, Berlin, Germany) administered with a 25 second delay.

5.2.2 Observer Performance

Three thoracic radiation oncologists specializing in treatment of NSCLC with SABR (observers 1–3) and three board-certified thoracic radiologists (observers 4–6), all of whom were blinded to outcomes, were asked to score all follow-up images as either benign injury/no recurrence or local recurrence. Data collection was completed using an in-house developed user interface in ClearCanvas Workstation 2.0 (Synaptive Medical, Toronto, Canada), as shown in Figure 5-1. All images were anonymized and randomized for each observer.
Figure 5-1: In-house developed user interface in ClearCanvas Workstation 2.0 (Synaptive Medical, Toronto, Canada) for data collection in observer study.

The custom-made interface in ClearCanvas was developed to mimic the clinical assessment of sequential scans over time. The interface first provided the physician with the planning CT scan to localize the pre-treatment lesion. After reviewing the planning scan, the first follow-up image was displayed. The physician was required to then provide an assessment (no recurrence vs. recurrence), and indicate the certainty level of the assessment on an ordinal scale (none, somewhat, moderately, very, or completely certain). Physicians were then asked to recommend the immediate next step for this patient (ignoring any non-local signs of disease); options included ongoing CT follow-up at standard or shorter time interval, biopsy, PET, or immediate salvage therapy. After entering these data, the responses were locked and the next follow-up image was shown. Assessments were made sequentially on all follow-up images available for each patient.
Observers were able to return to earlier scans in order to assess sequential changes, but could not modify responses. They also had full control over the interface to view and assess images at their own speed and adjust the window/level settings as needed. Observers were also provided written and verbal instruction prior to the reading session informing them of the task required, number of cases to be read and expected duration. To mitigate the assumption of a clinical prevalence of 10% local recurrences, observers were informed that the series of 45 patients contained 30–40% local recurrences. Of note, the term recurrence has been used throughout the chapter, although early assessment could indicate persistent disease.

At the first time point post-SABR, overall accuracy, false positive rate (FPR), and false negative rate (FNR) for physician assessment was measured. Sensitivity and specificity for assessing recurrence was also measured across all time points and available images. Fleiss' kappa was used to measure the reliability of agreement in all observations, across all six raters [20, 21]. Fleiss kappa is a measure of agreement between more than two raters giving categorical ratings, where agreement due to chance is factored out. A kappa less than or equal to 0 indicates poor agreement with agreement increasing as kappa approaches 1 (slight agreement=0.01-0.20, fair=0.21-0.40, moderate=0.41-0.60, substantial=0.61-0.80, and almost perfect agreement=0.81-1.00) [21]. Statistical analyses were completed in MATLAB 8.4 (The Mathworks, Natick, MA) with statistical testing at the 0.05 significance level.
5.2.3 Radiomic Analysis

Consolidative and peri-consolidative regions were delineated based on our semi-automated method, previously described in Chapter 4 [17, 18]. Briefly, on all follow-up CT images the tumour’s longest axial diameter based on *Response Evaluation Criteria in Solid Tumours* (RECIST 1.1) is measured during the normal clinical workflow and this is used to initialize the segmentation algorithm [22-24]. Further details pertaining to the region of interest delineation are outlined in Chapter 4. Examples of segmented consolidative and peri-consolidative regions are shown in Figure 5-2.

A custom radiomics and machine learning platform developed in-house was used to extract features from the consolidative and peri-consolidative regions using MATLAB 8.4. In total, 22 first-order features and 22 second-order grey-level co-occurrence matrix (GLCM) textures were calculated in both regions. The number of bins and density ranges for the GLCMs were set to yield 20 HU bin widths with densities between -1000 HU and 200 HU (60 bins) for the peri-consolidative regions and -600 HU and 200 HU (40 bins) for the consolidative regions. Due to the voxel anisotropy in post-SABR follow-up images, GLCMs were calculated within four in-plane neighbouring voxel directions [(-1,1,0), (-1,0,0), (0,1,0), (-1,-1,0)] within the 3D segmented volume, and texture features were averaged over all directions. Additionally, in the consolidative regions, 16 size- and shape-based features were calculated. These features were not calculated in the peri-consolidative regions as these regions’ sizes and shapes are defined according to the size and shape of the consolidation region and therefore would not be independent measurements. In total, 44 features were extracted from the peri-consolidative regions.
and 60 features from the consolidative regions, for a total of 104 features per image volume.

(1) Local recurrence

(2) Benign radiation induced lung injury

Figure 5-2: The 3-month post-SABR follow-up image for two patients who developed local recurrence (A and B) and two patients who developed only benign injury (C and D). The corresponding semi-automated consolidative (red) and peri-consolidative (green) segmentations are also shown.
Features were analyzed within the earliest follow-up time point post-SABR at 2–5 months, corresponding to the clinically scheduled 3-month follow-up scan. PRTools 5.0 (Delft Pattern Recognition Research, Delft, The Netherlands) was used for feature selection and classification [25]. Leave-one-out cross validation (LOOCV) was used for feature selection. At each fold of cross validation (CV), forward feature selection was performed on the training dataset only to select the optimal feature set, using the 1-nearest neighbour leave-one-out classification performance. A support vector machine (nu algorithm, linear kernel) was used to evaluate classification performance [26]. Nu is the regularization parameter which is the expected fraction of support vectors estimated by the leave-one-out-error of the 1-nearest neighbour rule. For a direct comparison to physician performance at 2–5 months post-SABR, classification error, FNRs and FPRs, where recurrence is defined as positive, were calculated. Across all folds of cross validation, a voting scheme was used to determine the most frequently selected features. These features were chosen as the radiomic signature. Evaluation on the entire dataset was completed by leave-one-out, 10-fold, 5-fold, and 3-fold CV to determine the impact of different training and testing set sizes on classification performance. The AUC was also determined for this dataset.

5.3 Results

5.3.1 Patient Baseline Characteristics

Baseline characteristics for all 45 patients are presented in Table 5-1. In total, 182 follow-up images were available for analysis with a median imaging follow-up of 20 months. The recurrence and non-recurrence patient groups were well matched, with a median age of diagnosis of 70 years (range 59–84) and median PTV size of 38 cm$^3$. 

162
(range 4–158 cm³). The fractionation scheme, and proportion of patients with central versus peripheral tumours, was also consistent across the groups.

**Table 5-1: Baseline patient characteristics.**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>All Patients (n = 45)</th>
<th>Recurrence (n = 15)</th>
<th>No-Recurrence (n = 30)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age – median (range)</td>
<td>70 (59–84)</td>
<td>71 (60–84)</td>
<td>70 (59–84)</td>
</tr>
<tr>
<td>Tumour size (mm) – mean (SD)</td>
<td>34 (15)</td>
<td>35 (14)</td>
<td>33 (15)</td>
</tr>
<tr>
<td>Male – N (%)</td>
<td>32 (71)</td>
<td>11 (73)</td>
<td>21 (70)</td>
</tr>
<tr>
<td>Charlson score – mean (SD)</td>
<td>2.7 (1.8)</td>
<td>2.9 (1.3)</td>
<td>2.7 (2.0)</td>
</tr>
<tr>
<td>Involved lobe – N (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LUL</td>
<td>15 (33)</td>
<td>4 (27)</td>
<td>11 (37)</td>
</tr>
<tr>
<td>LLL</td>
<td>11 (24)</td>
<td>4 (27)</td>
<td>7 (23)</td>
</tr>
<tr>
<td>RUL</td>
<td>11 (24)</td>
<td>3 (20)</td>
<td>8 (27)</td>
</tr>
<tr>
<td>RML</td>
<td>2 (4)</td>
<td>2 (13)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>RLL</td>
<td>6 (13)</td>
<td>2 (13)</td>
<td>4 (13)</td>
</tr>
<tr>
<td>Location – N (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central</td>
<td>6 (13)</td>
<td>2 (13)</td>
<td>4 (13)</td>
</tr>
<tr>
<td>Peripheral</td>
<td>39 (87)</td>
<td>13 (87)</td>
<td>26 (87)</td>
</tr>
<tr>
<td>PTV (cm³) – median (min, max)</td>
<td>38 (4, 158)</td>
<td>38 (5, 144)</td>
<td>40 (4, 158)</td>
</tr>
<tr>
<td>RT technique – N (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fixed Beam</td>
<td>37 (82)</td>
<td>12 (80)</td>
<td>25 (83)</td>
</tr>
<tr>
<td>Rapid Arc</td>
<td>8 (18)</td>
<td>3 (20)</td>
<td>5 (17)</td>
</tr>
<tr>
<td>Fractionation – N (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>12 (27)</td>
<td>4 (27)</td>
<td>8 (27)</td>
</tr>
<tr>
<td>5</td>
<td>21 (47)</td>
<td>7 (47)</td>
<td>14 (47)</td>
</tr>
<tr>
<td>8</td>
<td>12 (27)</td>
<td>4 (27)</td>
<td>8 (27)</td>
</tr>
</tbody>
</table>
5.3.2 Observer Performance

To determine the physicians’ ability to assess response over the entire course of follow-up, the sensitivity and specificity for identifying a recurrence at any time point during follow-up was assessed. Physicians had varying sensitivities and specificities for making a diagnosis of recurrence as observed in Table 5-2. The median sensitivity for physician assessment was 83.8% (range 67–100%) and median specificity was 75.0% (range 67–87%) across all observers. There was only a moderate level of agreement among all 6 observers, with a kappa value of 0.54 for all 182 assessments of recurrence vs. no recurrence. In all cases, radiologists (observers 4–6) had lower specificity compared to radiation oncologists (observers 1–3), but on average radiologists had higher sensitivity in detecting recurrence. The average time for each physician to first detect a local recurrence correctly post-SABR was greater than one-year post-SABR. However, radiologists were typically able to detect the recurrence earlier (mean of 13.4 months) than the radiation oncologists (mean of 18.2 months).

Although observers were informed that the dataset contained 30–40% local recurrences, the actual percentages of patients identified as recurrences varied among the observers. All three radiation oncologists assessed 38% of the patients as having a recurrence, consistent with the true percentage they were provided. However, all three radiologists assessed 49–53% patients as having a recurrence; a higher percentage than what was provided as truth.

The physicians’ certainty levels in their assessment varied substantially, with almost no agreement between observers (Table 5-2). One radiation oncologist (observer 2) and the three radiologists (observer 4-6) tended to score the majority of images as very
or completely certain. Across all images, physicians’ certainty of assessment had a Fleiss kappa of 0.06, indicating only slight agreement.

Table 5-2: Physician assessment of response post-SABR based on all follow-up images.

<table>
<thead>
<tr>
<th></th>
<th>Radiation Oncologists</th>
<th>Radiologists</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Observer 1</td>
<td>Observer 2</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>86.7%</td>
<td>80.0%</td>
</tr>
<tr>
<td>Specificity</td>
<td>86.7%</td>
<td>83.3%</td>
</tr>
<tr>
<td>Average Time to</td>
<td>16.8</td>
<td>19.4</td>
</tr>
<tr>
<td>Correctly Identify</td>
<td>(10.6)</td>
<td>(9.9)</td>
</tr>
<tr>
<td>Recurrences*, Months</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(SD)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Certainty Level,</td>
<td>None</td>
<td>Somewhat</td>
</tr>
<tr>
<td>Number of Images (%)</td>
<td>7</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>(3.8)</td>
<td>(18.7)</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>(3.8)</td>
<td>(10.4)</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>(0)</td>
<td>(26.9)</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>(0)</td>
<td>(14.3)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>(1.1)</td>
<td>(6.6)</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0)</td>
<td></td>
</tr>
</tbody>
</table>

*For each observer, the average time across all patients to first detect a true local recurrence correctly. Abbreviations: SD: standard deviation.

Figure 5-3 depicts the recommendations made for the next clinical intervention for each image dataset. Approximately 5% of all non-recurrence images (17 of 30 patients) were recommended for additional intervention with PET imaging. There were
also 8 patients in which at least one observer recommended more invasive interventions including biopsy or immediate salvage for patients with only benign injury. The interobserver agreement in terms of recommendation across all images had fair agreement with a Fleiss kappa of 0.31.

**Figure 5-3:** Inter-observer variability in follow-up recommendations across all follow-up image for patients with (REC) and without (NOREC) local recurrence.

The ability of physicians to predict recurrence within six months of SABR (using the 2–5 month follow-up scans) is shown in Table 5-3. At 2–5 months post-SABR, 38 of 45 patients had a CT image available for analysis. Physicians assessed the majority of images this time point as benign injury/no recurrence, with a false negative rate of 100% for five of six physicians. In 5 instances during this time period (of 228 total assessments), a PET scan was recommended. In no scenarios was a biopsy or surgical salvage recommended.
Table 5-3: Leave-one-out cross-validation results of the radiomic signature compared to physician assessment at 2–5 months post-SABR.

<table>
<thead>
<tr>
<th>Error</th>
<th>Radiation Oncologists</th>
<th>Radiologists</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Observer 1</td>
<td>Observer 2</td>
</tr>
<tr>
<td>Error</td>
<td>23.7%</td>
<td>34.2%</td>
</tr>
<tr>
<td>FPR</td>
<td>24.0%</td>
<td>0%</td>
</tr>
<tr>
<td>FNR</td>
<td>23.1%</td>
<td>100%</td>
</tr>
<tr>
<td>Percent</td>
<td>Very or</td>
<td>Certain</td>
</tr>
</tbody>
</table>

5.3.3 Radiomics Performance

To determine the final radiomic feature set for validation, the top features which had votes in at least 80% of the feature selection folds (i.e. at least 30 votes) were chosen. The top five most-selected features over CV folds were four features in the peri-consolidative region (minimum grey-level, GLCM homogeneity, GLCM correlation, and GLCM energy) and one feature in the consolidative region (grey-level uniformity) as shown in Table 5-4. This left a total of five features in the radiomic signature. Interestingly, nine of the top ten features were all appearance features within the peri-consolidative region.

To assess the predictive power of this five-feature radiomic signature, folded CV was performed. Leave-one-out CV produced an error of 23.7%, FPR of 24.0%, and FNR of 23.1%. The results of leave-one-out, 10-fold, 5-fold, and 3-fold cross validation are
shown in Figure 5-4. The radiomic signature demonstrated robustness to differences in training and testing dataset size, with classification error increasing by only 8% from leave-one-out to 3-fold CV. It also demonstrated balanced FPRs and FNRs, with the exception of 3-fold CV. The AUC for this radiomic signature on this dataset was 0.85. Examples of six correctly classified recurrence and injury cases in the LOOCV are shown in Figure 5-5. A false positive and false negative case are also shown for comparison.
Table 5-4: Top selected features at 2–5 months post-SABR based on leave-one-out feature selection (38 folds).

<table>
<thead>
<tr>
<th>Features</th>
<th>Equation</th>
<th>Region of Interest</th>
<th>Number of votes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum grey-level</td>
<td>The minimum intensity value of the region of interest.</td>
<td>Peri-consolidative</td>
<td>32</td>
</tr>
<tr>
<td>GLCM homogeneity</td>
<td>[ \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{P(i,j)}{1 +</td>
<td>i - j</td>
<td>} ]</td>
</tr>
<tr>
<td>GLCM correlation</td>
<td>[ \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{(i - \mu_i)(j - \mu_j)p(i,j)}{\sigma_i \sigma_j} ]</td>
<td>Peri-consolidative</td>
<td>31</td>
</tr>
</tbody>
</table>

Where:

\[ \mu_i = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} i P(i,j) \]

\[ \mu_j = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} j P(i,j) \]

\[ \sigma_i^2 = \sum_{i=1}^{N_g} (i - \mu_i)^2 P(i,j) \]

\[ \sigma_j^2 = \sum_{j=1}^{N_g} (j - \mu_j)^2 P(i,j) \]

GLCM energy

\[ \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P(i,j)^2 \]

Grey-level uniformity

\[ \sum_{x=1}^{N_t} P(x)^2 \]

Consolidative 30
*Grey-level co-occurrence matrix (GLCM) features: \( P(i, j, \) ) is the GLCM, \( N_g \) is the number of discrete grey level intensities, \( i \) is the \( i^{th} \) row of the GLCM, \( j \) is the \( j^{th} \) column of the GLCM.

First-order features: \( P(x) \) is the intensity value of voxel \( x \) within the region of interest; \( x = 1, 2, 3, \ldots, N \), where \( N \) is the number of voxels within the region of interest.

**Figure 5-4:** Results of folded cross-validation at 2–5 months post-SABR using the five-feature radiomic signature.
Figure 5-5: Example 2–5 month post-SABR scans for four patients who developed local recurrence (A-D) and four who developed only benign injury (E-H). Correctly classified recurrence (A-C) and injury patients (E-G) in LOOCV by the radiomic signature are shown. A false negative (D) and false positive (H) patient are also shown.

5.4 Discussion

Surveillance following SABR for lung cancer is a major clinical challenge as radiographic fibrosis can mimic local recurrence. Recent reports of patients undergoing ‘salvage resection’ for benign disease [4-6] underscore the difficulty in accurately distinguishing recurrence from fibrosis. To our knowledge, this is the first study assessing physician performance and accuracy in detecting local recurrence post-SABR. Overall, physicians demonstrated high sensitivity (83%) and moderate specificity (75%) for detecting recurrences, but the median time to detection of recurrence was 15.5 months. In addition, 9.4% of scans in patients with only benign changes resulted in a recommendation for further intervention with PET, biopsy, or immediate salvage treatment. Physicians were generally unable to detect recurrences within 6 months of
SABR. In contrast, radiomics assessment was able to correctly classify 76% of patients using scans done <6 months post-SABR.

In our study, scans performed within 6 months post-SABR did not often result in suspicion of local recurrence by physicians, and this finding is consistent with the published literature on early post-SABR imaging. In a study examining the impact of early post-SABR imaging within 6-months post-treatment in 62 patients, only 3% of patients had changes on those scans that ultimately led to a diagnosis of early recurrence [27]. Overall, we demonstrated the ability of physicians to detect recurrences at a later time point, typically when the HRF’s become more salient on imaging. Agreement between observers in terms of assessments of recurrence was only moderate and the certainly level of assessments varied across all images and physicians. This suggests the importance of a multi-disciplinary group discussion when assessing responses post-SABR.

The appearances detected by radiomics may be early indicators of the disease persistence and progression to local recurrence, which were not typically considered by physicians. Radiomics could augment a physician’s assessment of response post-SABR to potentially allow for earlier detection of recurrence post-SABR and fewer investigations of only benign fibrosis. Future work includes measuring the performance of the physician when provided with the output of the radiomics system. The importance of radiomic appearance features in the peri-consolidative region supports the work presented in Chapters 2-4 on the importance of features surrounding the consolidative regions for recurrence prediction [16-18]. As seen in Figure 5-5, patients with recurrence tend to have increased presence of ground-glass opacity surrounding consolidative
changes compared to patients with benign injury. The false negative cases predicted by radiomics tended to have minimal ground-glass opacity as seen in Figure 5-5(D) and false positives tended to have a larger presence of ground-glass opacity as seen in Figure 5-5(H). These appearances were similar to those cases correctly classified by the radiomics signature. Previous studies have also demonstrated the ability of CT texture analysis for predicting the development of radiation pneumonitis [28]. Radiomic signatures may identify patients in whom more frequent imaging, and regular reviews by multidisciplinary tumour boards are warranted.

This study must be considered within the context of its strengths and limitations. We used actual clinical cases with standard-of-care imaging follow-up, assessors were blinded, and we used a large number (six) of expert observers. We used solely CT imaging for physician assessment, however in clinical practice the patient’s clinical information would be available and may be considered in their assessment, and additional investigations, such as PET, could be ordered before proceeding to the next time point. In this study, although physicians were able to assess images on their own time with their preferred reading and room conditions, they were unable to obtain multidisciplinary input. In addition, not all patients with recurrence had pathologic confirmation, as pathologic confirmation of recurrence may not be possible in patients who are ill due to comorbidities, or may not be pursued if no treatment options are available. Also, since histologic patterns of recurrence post-SABR are still unknown, validation of our radiomic signature with histology is needed. Future work includes the assessment of physician and radiomic performance in the context of pathologic specimens for lung SABR patients planned for surgical resection within 10 weeks of radiotherapy (MISSILE clinical trial,
NCT02136355). This will involve correlating post-SABR CT changes with histologic findings.

Regarding the radiomic signature, analysis was only performed at the early 2–5 month time point. Future studies should be completed on how radiomic performance changes as time post-SABR increases. Although this is one of the largest series of recurrences post-SABR, further validation of the radiomics signatures is needed on a larger dataset. Another limitation of the current study is the use of a pre-validated radiomic signature for comparison to physicians’. Future work includes measuring radiomic performance on a separate validation set as well as prospective studies.

5.5 Conclusions

Physician assessment of post-SABR CT images shows high sensitivity and good specificity, but median time to detection of recurrence is long, inter-observer agreement is moderate, and unnecessary interventions are sometimes recommended for patients with only benign changes. Although physicians generally perform poorly in detecting recurrences within 6 months post-treatment, radiomic assessment demonstrates excellent classification ability during that time point, with an AUC of 0.85. This suggests that there are CT features predictive of recurrence that are not salient to the human expert at this early, clinically-important time point. This radiomic tool has the potential to lead to a clinically useful computer-aided decision support tool based on routinely-acquired CT imaging, which could lead to earlier salvage opportunities and fewer unnecessary invasive investigations of patients with only benign injury.
5.6 References


Chapter 6

6 Conclusions and Future Directions

The final chapter of this thesis reexamines the research objectives and provides a summary of the key findings and conclusions of Chapters 2-5. This chapter will also address limitations of the current study and potential solutions. Finally, future directions for response assessment following SABR are presented.

6.1 Overview of Rationale and Research Objectives

Stereotactic ablative radiotherapy (SABR) is now a guideline recommended treatment option for patients with early-stage NSCLC [1]. SABR has demonstrated local control rate similar to surgery of over 90% at 3 years follow-up [2, 3]. However, the presence of substantial post-radiotherapy radiological changes are found on CT follow-up, which can appear with similar shape and size to a recurring tumour [4]. Misclassification of benign fibrosis as local tumour recurrence has been reported, resulting in patients undergoing unnecessary and invasive surgical procedures to remove only scar tissue [5]. Current approaches to aid in response assessment have focused on qualitative high-risk features on CT imaging or metabolic imaging with FDG-PET, which do not typically detect recurrence until after a year following treatment. There is currently an unmet clinical need to provide timely and accurate assessment of response following SABR for early stage NSCLC.

The recent introduction of radiomics, has demonstrated the ability to quantify tumour phenotypes based on quantitative image features [6, 7]. This has the potential to personalize patient care based on predicted responses to therapy and has the ability to detect treatment failure at an earlier time point following treatment. In response to this
unmet clinical need, we aim to develop a radiomics decision support tool to aid in response assessment following SABR for early stage NSCLC. A means for predicting recurrence on routine CT imaging within 6 months of treatment would allow for timely intervention of recurrence, which typically manifests after 1 year [8]. Accordingly, the overarching objective of this thesis was to develop a radiomic software system to aid in treatment response assessment on follow-up CT after SABR.

6.2 Summary and Conclusions

In this thesis the overarching objective was evaluated by testing the following central hypothesis: a radiomic software system will outperform physicians in the early assessment of response post-SABR. Radiomics demonstrated the ability to predict local recurrence with higher accuracy than physicians within 6 months of treatment. This thesis has advanced knowledge and technology towards the improvement of response assessment following SABR for early stage NSCLC. Prior to this work, there was no study assessing a physician’s performance in distinguishing fibrosis from recurrence following SABR for lung cancer. The use of quantitative image appearance features to detect recurrence following SABR had also never been investigated.

Prior to the work presented in this thesis, physicians relied on qualitative image characteristics of an enlarging opacity and the RECIST guidelines for basic quantitative measurements on CT imaging, to assess response after SABR. PET imaging is also used to detect recurrence after SABR; however most of these tools do not typically detect recurrence until a year after treatment. The use of biopsy to confirm recurrence is also used, but can be risky and may not be feasible in all patients. The work completed in this thesis has advanced knowledge and technology, suggesting the possibility for integration
of quantitative radiomic features to assist physicians in early detection of recurrence following SABR. Quantitative image features can be easily acquired on routine follow-up CT imaging after SABR. These features also have the ability to obtain much richer information from the CT image compared to a single measurement which is typically done with RECIST. Compared to response assessment on PET imaging, CT is more accessible, is inexpensive, does not require a radiotracer or additional imaging session, and standardization is less complex. An automated decision support system may also be more reproducible and consistent compared to qualitative image assessment, which can be subject to operator variability. CT is also less invasive and risky compared to a biopsy. This decision support system outperformed physicians for the early assessment of response and has the potential to improve response assessment following SABR by providing more accurate and timely prediction of recurrence compared to standard techniques in the current clinical workflow.

In Chapter 2, we evaluated the potential of quantitative image feature analysis for distinguishing recurrence and benign fibrosis patient groups. In this hypothesis generating study, regions of common post-SABR consolidative and ground-glass opacity changes were manually delineated on all follow-up CT images for 24 lesions treated with SABR. Quantitative size and appearance image features were extracted from both regions and analyzed according to time post-SABR. Traditional size measurements including longest axial diameter and 3D volume could not distinguish patient groups until 15 months post-SABR. Appearance features including density of consolidative changes and first-order texture of ground-glass opacity regions could distinguish patient groups at 9 months post-SABR. These findings suggest that advanced appearance measures may
provide an earlier detection of recurrence compared to traditional measure of size. Thus, the thesis advanced knowledge in SABR response assessment by demonstrating that quantitative appearance features of post-SABR CT changes were able to significantly distinguish local recurrence and benign injury patient groups.

In Chapter 3, we examined the ability of radiomic features to predict recurrence in individual patients. Second-order texture features were evaluated for their ability to predict local recurrence within 6 months post-SABR and compared to traditional measures of size. Once again, manual delineation of post-SABR consolidation and ground-glass opacity was performed. We demonstrated the ability of second-order texture features within regions of ground-glass opacity to predict recurrence within 6 months post-SABR with accuracies greater than 75%. Traditional size measures at the same time point produced accuracies less than 60%. Classification based on the clinical standard response evaluation criteria in solid tumours (RECIST) progressive disease guidelines, demonstrated accuracies of 52%. These findings suggest that CT texture features on post-treatment images taken within 6 months post-SABR has the potential to predict eventual local recurrence. Thus, the thesis advanced knowledge in SABR response assessment by demonstrating that texture features within post-SABR regions of GGO can predict eventual cancer recurrence in individual patients within 6 months post-SABR with higher accuracies than standard measures of size and response criteria measures.

In Chapter 3, we also examined the impact of perturbations on the borders of the manually segmented regions of post-SABR changes. We demonstrated minimal changes in classification performance with perturbations of the manual GGO boundaries, signifying that perhaps a complete and accurate delineation of the GGO may not be
necessary for prediction of recurrence with radiomic features. *Thus, the thesis advanced knowledge in SABR response assessment by demonstrating that complete delineations of the GGO regions are not needed for recurrence prediction, suggesting the use of a sampling method could allow for a faster segmentation and clinical translation of a radiomics decision support tool to aid in assessing response.*

In Chapter 4, we developed and evaluated the use of a graph cuts based semi-automated segmentation system to predict local recurrence following SABR. For clinical translation of our system to a useful computer-aided diagnosis system, the use of a faster and more reproducible delineation or sampling method of the GGO regions is needed. A semi-automated segmentation system was developed using the clinical standard longest axial diameter measurement taken as part of the RECIST guidelines. This system segmented the consolidative changes and performed a 3D concentric expansion to create a peri-consolidative region, intended to sample regions of GGO surrounding the consolidative mass. Radiomic features could then be extracted for classification. This system was evaluated with three expert observers to examine consistency and reproducibility of segmentations and classification performance. We found that our semi-automated system produced more reproducible segmentations and could predict recurrence with non-inferiority to manual segmentations. *Thus, the thesis advanced knowledge in SABR response assessment by demonstrating a semi-automated decision support system could predict recurrence with similar accuracies to a manual approach.* *This system is an advance in technology that was 20 times faster than manual delineations and could be integrated into the clinical workflow to aid in response assessment post-SABR.*
In Chapter 5, to provide a basis of comparison for our radiomics system, an observer study to measure expert physician performance in response assessment was completed. Three radiation oncologists and three radiologists scored all follow-up images after SABR and classified the images as recurrence or no recurrence. Physicians had high sensitivity (median of 83%) and specificity (median of 75%) for detecting recurrence at any time point during follow-up. However, the median time to detect local recurrences was 16 months following treatment. Radiologists tended to have higher sensitivity but lower specificity for detecting recurrence compared to radiation oncologists. Within 6 months of treatment, physicians tended to assess all images as no recurrence, suggesting there are not suspicious of recurrence at this early time point and tend to wait until an enlarging mass is evident on CT to assess patients as a local recurrence. Thus, the thesis advanced knowledge in SABR response assessment by demonstrating a physician’s ability to detect local recurrence on follow-up CT imaging and provides for a base level of performance to assess the incremental value of a decision support system.

Also in Chapter 5, a multi-feature radiomic signature was developed on a larger dataset of 45 patients to predict recurrence within 6 months of treatment. We identified five radiomic appearance features within regions of post-SABR consolidative and peri-consolidative regions which could predict recurrence with a leave-one-out cross validation error of 23%. Texture features within the peri-consolidative regions were the top performing features, demonstrating that appearance measures within these regions are more predictive of recurrence compared to size, shape, or appearance features in the consolidative regions. Patients with recurrence tended to have increased presence of GGO surrounding the consolidative changes. The appearances in the GGO also tended to
be spiculated and were emerging from the consolidative mass, compared to a smoother appearance in patients with only benign fibrosis. Radiomics outperformed physicians at this early follow-up time point, where physician assessed almost all images as no recurrence. These appearances picked up by radiomics may be early indicators of either disease persistence following treatment or progression to local recurrence. *Thus, the thesis advanced knowledge in SABR response assessment by demonstrating the ability of radiomic appearance features of post-SABR radiological changes to predict eventual cancer recurrence earlier than expert physicians. This demonstrates the possible role of radiomic features to provide additional information for physicians by means of a decision support tool to assist in response assessment post-SABR.*

In summary, we have provided: (1) evidence that quantitative image appearance features of post-SABR radiological changes can distinguish recurrence and benign patient groups earlier than traditional measures of size; (2) evidence that radiomic texture features within post-SABR ground-glass opacity changes can predict recurrence in individual patients within 6 months of treatment; (3) evidence that a new semi-automated segmentation system could delineate regions of common post-SABR changes based on the clinically acquired longest axial diameter measurement, extract radiomic features, and predict recurrence with non-inferiority to manual segmentations; and (4) evidence that radiomics can detect early changes within 6 months of SABR treatment that may be associate with disease persistence or progression to local recurrence which are typically not picked up by expert physicians.
6.3 Limitations

The contents of this thesis must be considered within the context of its limitations. Specific limitations for Chapters 2-5 are presented in the discussion section of the respective Chapters. However, there are general limitations of this thesis that should be addressed. One limitation to all studies was that a relatively small dataset of patients was analyzed. All patients analyzed throughout this thesis were treated and followed at the same institution. Therefore all follow-up imaging was performed at the same centre, with a standard acquisition and reconstruction protocol. Although images were from a single institution with set scanning parameters, variation may exist between scanners [9]. This system should be evaluated with a larger dataset of patients from multiple institutions, to account for any variability image acquisition or reconstruction. Another limitation of this work is that not all patients with recurrence had pathologic confirmation. However, to avoid imaging-based definitions of recurrence which may introduce substantial bias, in these studies, patients with recurrence were assessed based on multidisciplinary group discussion, a combination of PET and CT findings and eventual clinical outcomes.

6.4 Future Directions

The methods developed in this thesis support research initiatives in several future directions. This section will address remaining gaps in knowledge that exist and potential applications of this system on future research directions.

6.4.1 External Validation and Evaluation of Additional Imaging

Radiomics has shown the potential to outperform expert physicians in the early assessment of response following SABR treatment. However, prior to clinical validation,
radiomic features must be validated on multiple large independent datasets. We have shown the potential of radiomics to predict local cancer recurrence within 6 months of treatment. However, further investigation on radiomics at later follow-up time points is warranted and may provide supplemental information to the early time points. Delta radiomics is an emerging area of study which refers to analyzing the change in radiomic features over time. The use of delta radiomics to assess changes in radiomic features over the course of post-SABR follow-up may provide additional information in longitudinal image analysis [10]. A thorough investigation of pre-treatment tumour characteristics may also yield insight into factors associated with recurrence [11]. This could result in personalized management of lung cancer patients to inform treatment decision making or risk stratification. The combination of these approaches may provide additional information to improve response assessment following SABR and provide personalized patient care and follow-up assessment.

6.4.2 Incorporation of High-Level Domain Knowledge

This thesis has made significant contributions to improve timely and accurate assessment of response following SABR for early stage NSCLC. However, there remains room for improvement in SABR response assessment. To date our decision support system to aid in treatment response assessment has focused on analysis of the treatment area on follow-up CT. Our analysis has focused on evaluating regions of post-SABR consolidative and ground-glass opacity changes. Therefore, we have been focusing on solely the quantitative characteristics of the lesion on CT imaging. However, more information may be available regarding the lesion or patient as a whole. Therefore, the utility of additional features in combination with radiomics should be evaluated.
Qualitative high risk CT features have been independently validated and could provide useful information at later follow-up time points [12, 13].

The use of FDG-PET imaging has been shown to distinguish local recurrence from benign RILI on pre- and post-treatment imaging. Previous work has also demonstrated that abnormal texture of pre-treatment SUV uptake was associated with nonresponse and poorer prognosis following chemo-radiotherapy for NSCLC [14]. The combination of PET and CT features has also been shown to predict response in lung cancer [15]. The addition of quantitative SUV values on PET may provide additional metabolic information to complement the radiomic features seen on CT, and allow for a more comprehensive view of the tumour. Although there are limitations with using post-SABR SUV values for response assessment, the use of radiomic analysis on post-treatment PET imaging may provide a measure of regional heterogeneity of SUV uptake values within the tumour or reveal changes from pre-treatment values. This information may play a role in response following SABR and warrants further investigation. The combination of radiomic information on CT and PET can provide phenotypic and metabolic information regarding the tumor. However, the addition of genomic or molecular markers may provide associations of outcomes or radiomic features with gene expression patterns [6].

Although all patients in our study were early stage (T1/T2, N0) lung cancer, there can exist substantial variability in patient history and clinical characteristics of the disease. Tumour size has been shown to be an important predictor of local control following treatment for NSCLC [16]. Therefore, characteristics of the patients’ disease, including T stage, may augment the decision support system. Additional clinical
information and prognostic factors related to the patients’ history, including age, gender, past smoking history, lung function, and performance status could also be useful [17]. This thesis did also not consider the effects of co-morbidities and the potential impact on the normal lung tissue on CT. Future work analyzing the impact of these co-morbidities on normal lung appearances and radiomic features warrants further investigation.

Pathological features including tumour grade and histological cell type, as well as other genomic or biological marker may also play a role in lung cancer prognosis and should be considered in a decision support system [18]. These clinical prognostic factors may augment radiomic image features to improve prediction of local recurrence following SABR.

All patients in our study were treated with a risk-adapted approach with one of three fractionation regimens: 60 Gy in 3, 5, or 8 fractions. The impact of different dose fractionation regimes has not been considered in any of our studies and may provide additional information for predicting recurrence. Other dosimetric factors of both the tumour (e.g. maximum PTV dose) and normal tissue (e.g. V20, percent of the lung volume receiving 20 Gy or more) may be useful for predicting outcomes. The dose-response relationships may also provide additional information regarding the radiological changes seen post-SABR [19]. Registration of pre-treatment dose distributions to follow-up images may allow for the association of radiomic features with dose levels. This could provide insight into changes associated with high or low dose regions and the location of the local recurrence with respect to dose received.
6.4.3 Use in Clinical Decision Support

Currently assessment of response following SABR relies on expert qualitative assessment [20]. A combination of metabolic imaging with FDG-PET and/or biopsy may augment and confirm a physician’s assessment of response. The radiomic decision support tool developed in this thesis has the potential to aid in this clinical decision. Therefore, the impact of a radiomic decision support system on clinical management must also be determined in an additional observer study. First, effective presentation of this information from the decision support system must be investigated. The decision support tool could provide information on the regions of interest being analyzed or a confidence value of the classifier. Providing information regarding the radiomics confidence or decision could be presented in a continuous or discrete manner. Consultation with radiation oncologist and radiologist users should be performed to determine preferences and suggestions for decision support tools.

The integration of computer aided detection and diagnosis tools in lung cancer has demonstrated the ability to improve radiologists performance in detecting and diagnosing lung nodules [21]. The work completed in this thesis was the first-ever study to determine physician ability to assess response post-SABR. Effects of this decision support tool on physician assessment of response post-SABR must be assessed to determine the incremental value of performance without any decision support. Training of the operators may also have an impact on the effect of the decision support tool, and selection of physicians for this study must be considered [22]. Integration into the clinical workflow could allow for earlier detection of local recurrence following SABR to allow for timely salvage interventions. If this tool demonstrates success by improving a physician’s
assessment of response, a prospective clinical trial evaluating its impact on clinical decision making and patient outcomes could be determined.

Finally, the correlation of radiomic features with post-SABR histology would also be extremely valuable. This would allow for evaluation of tissue components within different regions of post-SABR changes. This could provide insight into the biological properties associated with certain radiomic features. This information may be important to provide to physician users to allow for the interpretation of what the radiomic features mean biologically.

6.5 References


Appendices

APPENDIX A – Permission for Reproduction of Scientific Articles

A.1 Permission to reproduce previously published material in Chapter 1

Permission form

Requestor Information:
Name: Sarah Mattonen
Affiliation: The University of Western Ontario
Address:
Email:

What Publication Material Would You Like To Use/Adapt?
DOI: 10.3978/j.issn.2072-1439.2013.11.09
Authors: Sarah A. Mattonen, Kitty Huang, Aaron D. Ward, Suresh Senan, David A. Palma
Article Title: New techniques for assessing response after hypofractionated radiotherapy for lung cancer
Title of Publication: Journal of Thoracic Disease
Page Numbers of Article: 375-386
Page Numbers of Material in Question: 375-386
Description (figure, table, article, etc.): Full article

Purpose of reuse:
Purpose of reuse (Educational, Commercial, etc.): PhD thesis (integrated-article)
Description: Chapters may include published/submitted/unpublished work
- Print copy (specific context, quantity, timeline, etc.):
- E-copy/database (specific context, quantity, timeline, URL, etc.):

Expected completion date: August 2016.

Permission is granted as stipulated with signature of both requestor and copyright holder's. The source of the article is required to be noted.

Date: April 4th, 2016
Signature (Requestor):

Date: June 14, 2016
Signature (Copyright holder):

AME Publishing
A.2  Permission to reproduce previously published material in Chapter 1
(http://www.birpublications.org/page/permissions#thesis)

Interested in permission to reuse content from this article?

Here are some things to consider:

- Is the original article open access?
- Are you the author of the original article?
- Do you want to reproduce content from the article in a PhD thesis?
- All other queries

Is the original article open access?

Open Access articles in BJR and DAMFR are accompanied by an open access statement on the title page, detailing the type of open access license.

All BJR-case reports articles are published on an open access basis under a Creative Commons attribution license (CC BY).

You can read more about our open access policy here:

- Creative Commons attribution non-commercial licence (CC BY NC)
  - If the article is open access under a Creative Commons attribution non-commercial licence (CC BY NC), you are free to reuse, change and build upon the work, even commercially. Your new work must be non-commercial and acknowledge the original source.
- Creative Commons attribution licence (CC BY)
  - If the article is open access under a Creative Commons attribution licence (CC BY), you are free to reuse, change and build upon the work, even commercially. Your new work must acknowledge the original source.

Are you the author of the original article?

- BJR, DAMFR and Imaging
  - Authors of BJR, DAMFR and Imaging, articles are free to reuse figures, tables, charts and short sections of text (up to 300 words) from their article in other publications, free of charge providing they cite the original source article.
- BJR case reports
  - All BJR-case reports articles are open access under a Creative Commons attribution licence (CC BY), so authors (and third parties) are free to reuse, change and build upon the work, even commercially. Your new work must acknowledge the original source.

Do you want to reproduce content from the article in a PhD thesis?

If you are the author of the original article and wish to reproduce the article in full or in part in your PhD thesis you are free to do so providing you cite the original source article.

If you are not the author of the original article but wish to reproduce 1-2 figures, tables, charts and short sections of text (up to 350 words) from a BJR, DAMFR or Imaging article, you are free to do so providing you cite the original source article. Requests to reproduce more than 2 figures, tables, charts or large sections of text must be processed by the copyright clearance center (see below).

All other queries

All other permission queries are handled by the copyright clearance center:

- Search for "British Journal of Radiology" or ISSN: 0007-128X, then
- Select "pay per view option" then
- "Purchase or display content": then
- you will then be able to select the parameters you require and will be taken through the permission application process.

If you still have questions, please contact publications@bir.org.uk.
A.3  Permission to reproduce previously published material in Chapter 2

RE: ionc20:Distinguishing radiation fibrosis from tumour recurrence after stereotactic ablative radiotherapy (SABR) for lung cancer: A quantitative analysis of CT density changes

Academic Journals Society Permissions <society.permissions@tandf.co.uk>

Wed 4/5/2015 9:51 AM

To: Sarah Matteson

Our Ref: MD/IONC/P6590

6th April 2016

Dear Sarah Matteson,

Thank you for your correspondence requesting permission to reproduce the following article published in our journal in your printed thesis and to be posted in your university’s repository at the University of Western Ontario.


We will be pleased to grant permission on the sole condition that you acknowledge the original source of publication and insert a reference to the article on the Journals website:

This is the authors accepted manuscript of an article published as the version of record in Acta Oncologica

Please note that this license does not allow you to post our content on any third party websites or repositories.

Thank you for your interest in our Journal.

Yours sincerely


Taylor & Francis is a trading name of Informa UK Limited, registered in England and Wales, no. 1072954

From: Sarah Matteson
Sent: 01 April 2016 22:12
To: Academic UK Non Rightlink
Subject: ionc20 Distinguishing radiation fibrosis from tumour recurrence after stereotactic ablative radiotherapy (SABR) for lung cancer: A quantitative analysis of CT density changes

Permissions Request
Contact name: Sarah Matteson

06/04/2016 8:57 AM
A.4  Permission to reproduce previously published material in Chapter 3

DATE OF REQUEST:  March 15, 2016

FROM:  Sarah Mattonen

EMAIL ADDRESS:

1. Permission is granted to:
   Sarah Mattonen

2. Permission is requested to use the following material:

3. For what purpose:
   As a chapter in my PhD thesis

Authors seeking permission must also notify the first author of the article from which permission is being sought.

Permission is hereby granted:  March 16, 2016

Date

Submit via Email  Reset Form

The Association’s Journals are Medical Physics and Journal of Applied Medical Physics
Member Society of the American Institute of Physics and the International Organization of Medical Physics

1631 Prince Street | Alexandria, VA 22314-2818 | phone 571.298.1300 | fax 571.298.1301 | www.aapm.org
A.5  Permission to reproduce previously published material in Chapter 4

RE: Request for Permission to Use Published Materials

Nicole Harris
Thu 3/26 4:29 PM

To: Sarah Ashley Mattomen

Dear Ms. Mattomen,

Thank you for seeking permission from SPIE to reprint material from our publications. As author, SPIE shares the copyright with you, so you retain the right to reproduce your paper in part or in whole. Publisher’s permission is hereby granted under the following conditions:

1. The material to be used has appeared in our publication without credit or acknowledgment to another source; and
2. You credit the original SPIE publication, include the authors’ names, title of paper, volume, issue, SPIE volume number, and year of publication in your credit statement.

Sincerely,

Nicole Harris
Administrative Editor, SPIE Publications
1100 29th St.
Bellingham, WA 98225
+1 360 676 5188 (office)

SPIE is the international society for optics and photonics. http://SPIE.org

From: Sarah Ashley Mattomen
Sat: Tuesday, March 15, 2016 10:55 AM
To: nicole@spie.org
Subject: Request for Permission to Use Published Materials

Good afternoon,

I would like to request permission to use the following journal article as a chapter in my PhD thesis:


Please let me know if you require any further information.

Best regards,

Sarah Mattomen
A.6  Permission to reproduce previously published material in Chapter 5

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction and that they are available at any time at http://righlink.elsevier.com.

GENERAL TERMS
The following terms and conditions apply to specific license types:

15. Translation: This permission is granted for non-exclusive world English rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article.

16. Posting licensed content on any Website: The following terms and conditions apply as follows: Licensing material from an Elsevier journal: All content posted to the web site must maintain the copyright information line on the bottom of each image. A hyper-text must be included to the HomePage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com. Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEra.

Licensing material from an Elsevier book: A hyper-text link must be included to the Elsevier homepage at http://www.elsevier.com. All content posted to the web site must maintain the copyright information line on the bottom of each image.

Posting licensed content on Electronic reserve: In addition to the above the following clauses are applicable: The web site must be password-protected and made available only to bona fide students registered on a relevant course. This permission is granted for 1 year only. You may obtain a new license for future website posting.

17. For journal authors: the following clauses are applicable in addition to the above:

Preprints:
A preprint is an author's own write-up of research results and analysis, it has not been peer-reviewed, nor has it had any other value added to it by a publisher (such as formatting, copyright, technical enhancement etc.). Authors can share their preprints anywhere at any time. Preprints should not be added to or enhanced in any way in order to appear more like, or to substitute for, the final versions of articles however authors can update their preprints on arXiv or BioPBeP with their Accepted Author Manuscript (see below).

If accepted for publication, we encourage authors to link from the preprint to their formal publication via its DOI. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help users to find, access, cite and use the best available version. Please note that Cell Press, The Lancet and some society-owned have different preprint policies. Information on these policies is available on the journal homepage.

Accepted Author Manuscripts: An accepted author manuscript is the manuscript of an article that has been accepted for publication and which typically includes author-incorporated changes suggested during submission, peer review and editor-author communications.

Authors can share their accepted author manuscript:

- immediately
  - via their non-commercial person homepage or blog
  - via their research institute or institutional repository for internal institutional users or as part of an invitation-only research collaboration work-group
  - directly by providing copies to their students or to research collaboration for their personal use
  - for private scholarly sharing as part of an invitation-only work group on commercial sites with which Elsevier has an agreement
- after the embargo period
  - via non-commercial hosting platforms such as their institutional repository
  - via commercial sites with which Elsevier has an agreement

In all cases accepted manuscripts should:

- link to the formal publication via its DOI
- bear a CC-BY-NC-ND license - this is easy to do
- if aggregated with other manuscripts, for example in a repository or other site, be shared in alignment with our hosting policy not to be added to or enhanced in any way to appear more like, or to substitute for, the published journal article.

Published journal article (PJA): A published journal article (PJA) is the definitive final record of published research that appears or will appear in the journal and embodies all value-adding publishing activities including peer review co-ordination, copy-editing,
formatting, (if relevant) pagination and online enrichment.

Policies for sharing publishing journal articles differ for subscription and gold open access articles.

**Subscription Articles:** If you are an author, please share a link to your article rather than the full-text. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help your users to find, access, cite, and use the best available version.

There and dissertations which contain embedded PIAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

If you are affiliated with a library that subscribes to ScienceDirect you have additional private sharing rights for others' research accessed under that agreement. This includes use for classroom teaching and internal training at the institution (including use in course packs and coursework programs), and inclusion of the article for grant funding purposes.

**Gold Open Access Articles:** May be shared according to the author-selected end-user license and should contain a CrossMark logo, the end user license, and a DOI link to the formal publication on ScienceDirect.

Please refer to Elsevier's posting policy for further information.

18. For book authors the following clause are applicable in addition to the above:

Authors are permitted to place a brief summary of their work online only. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version. Posting to a repository: Authors are permitted to post a summary of their chapter only in their institution's repository.

19. Thesis/Dissertation: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for ProQuest/UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission. Theses and dissertations which contain embedded PIAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

**Elsevier Open Access Terms and Conditions:**

You can publish open access with Elsevier in hundreds of open access journals or in nearly 2000 established subscription journals that support open access publishing. Permitted third party re-use of these open access articles is defined by the author's choice of Creative Commons user license. See our open access license policy for more information.

**Terms & Conditions applicable to all Open Access articles published with Elsevier:**

Any reuse of the article must not represent the author as endorsing the adaptation of the article nor should the article be modified in such a way as to damage the author's honour or reputation. If any changes have been made, such changes must be clearly indicated. The author(s) must be appropriately credited and we ask that you include the end user license and a DOI link to the formal publication on ScienceDirect.

If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source it is the responsibility of the user to ensure their reuse complies with the terms and conditions determined by the rights holder.

**Additional Terms & Conditions applicable to each Creative Commons user license:**

**CC BY:** The CC-BY license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article and to make commercial use of the Article (including reuse and/or resale of the Article by commercial entities), provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. The full details of the license are available at [http://creativecommons.org/licenses/by/4.0](http://creativecommons.org/licenses/by/4.0).

**CC BY NC SA:** The CC-BY-NC-SA license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article, provided this is not done for commercial purposes, and that the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. Further, any new works must be made available on the same conditions. The full details of the license are available at [http://creativecommons.org/licenses/by-nc-sa/4.0](http://creativecommons.org/licenses/by-nc-sa/4.0).

**CC BY NC ND:** The CC-BY-NC-ND license allows users to copy and distribute the Article, provided this is not done for commercial purposes and further does not permit distribution of the Article in any way, and provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, and that the licensor is not represented as endorsing the use made of the work. The
full details of the license are available at http://creativecommons.org/licenses/by-nc-nd/4.0

Any commercial reuse of Open Access articles published with a CC BY-NC-SA or CC BY-NC-ND license requires permission from Elsevier and will be subject to a fee.

Commercial reuse includes:
- Associating advertising with the full text of the Article
- Charging fees for document delivery or access
- Article aggregation
- Systematic distribution via e-mail lists or share buttons

Posting or linking by commercial companies for use by customers of those companies.

20. Other Conditions:

v1.8

Questions? customerservice@copyright.com or +1-855-239-3415 (toll free in the US) or +1-416-640-2777.
APPENDIX B – Supplementary Material to Chapter 2

Scan Parameters
Planning scans used for treatment planning were acquired using four dimensional (4D) CT (Lightspeed 16; GE Medical Systems, Waukesha, USA) at 140 kVp, 100-110 mAs, and a 2.5 mm slice thickness. 4D CT was completed using free breathing with binning by respiratory phase. Pre-treatment 4D-CTs used for image analysis were maximum inspiration of each patient.

For the purposes of this analysis, diagnostic CT scans taken post-treatment were used. All diagnostic CT scans done at the treating center were performed on one of three scanners [Siemens Volume Zoom 4-slice, Siemens Sensations 64-slice (Siemens Nederland N.V., Den Haag, Netherlands) or Philips Brilliance iCT 256-slice (Royal Philips Electronics, Inc., Amsterdam, Netherlands)]. Standard machine settings for post-treatment scans were 120 kVp, 100 mAs, spiral acquisition, 0.5 second rotation time, and 2.5-5 mm slice thickness. Scans were acquired at inspiratory breath hold and 70 cc of intravenous contrast [Ultravist-300; Bayer Pharma AG, Berlin, Germany] was administered with a delay of 25 seconds.

Importing and Contouring
All images were received in an anonymized form and imported into ClearCanvas Workstation 2.0 (ClearCanvas Inc., Toronto, Canada). Contouring was completed in ITK-SNAP (Version 2.2.0) (15) using a lung window setting (window width of 1500 HU and window level of -600 HU), and a mediastinal window setting (window width of 350 HU and window level of 40 HU) for tumours or fibrosis abutting the mediastinum. Contours were manually segmented slice-by-slice using a polygon tool and/or paintbrush tool.

Calculation of Measures
Response Evaluation Criteria in Solid Tumours (RECIST) measures were taken using the ClearCanvas Workstation ruler, for the consolidative areas on the pre- and post-treatment images, according to RECIST 1.1. MATLAB 7.13 (The MathWorks Inc., Natick, MA, USA) was used in the calculation of three dimensional (3D) volume and CT density measures for both the consolidative and ground glass opacity contours. 3D volume was
calculated by determining the voxel volume encompassed by the contours, based on the voxel dimensions. The number of voxels contained within each contoured region was determined and then multiplied by the voxel volume to determine 3D volume of each contoured region. CT density was determined based on the mean intensity values (HU) of all voxels in each contoured region of interest.

**Analysis of Measures**

The measures were analysed cumulatively from the treatment end date until a specific follow-up time point. The follow-up time points used in this study included 3, 6, 9, 12, 15, 18, 21, 24, 36, and 42 months post treatment. For a given time point, all images up to that time point were used in the analysis. For a given lesion, the following cumulative measures were calculated:

1) **RECIST** – Mean of all RECIST measures per scan
2) **CT density** – Mean of all voxels’ intensities (HU)
3) **Standard deviation of CT density** – Standard deviation of all voxel intensities (HU)
4) **3D volume** – Mean of all 3D volume measures per scan

Differences between groups were assessed using an independent samples t-test with unequal variances in MATLAB 7.13 (The MathWorks Inc., Natick, MA, USA). All statistical tests were two-sided with $p \leq 0.05$ indicative of statistical significance.
**APPENDIX C – Supplementary Material to Chapter 4**

**Table C1:** Individual texture feature values and classification metrics including error, false positive rate (FPR), false negative rate (FNR), and area under the receiver operating characteristic curve (AUC) at 2–5 months post-SABR. All values are reported as the mean ± the standard deviation.

<table>
<thead>
<tr>
<th></th>
<th>Observer</th>
<th>Feature Value - Recurrence</th>
<th>Feature Value - Injury</th>
<th>p-value</th>
<th>Error</th>
<th>FPR</th>
<th>FNR</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Correlation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operator 1 (Manual)</td>
<td>0.006 ± 0.002</td>
<td>0.01 ± 0.007</td>
<td>0.098</td>
<td>33.6 ± 10.9</td>
<td>37.8 ± 12.9</td>
<td>29.0 ± 12.8</td>
<td>0.61 ± 0.06</td>
<td></td>
</tr>
<tr>
<td>Operator 1 (Semi-auto)</td>
<td>0.008 ± 0.001</td>
<td>0.014 ± 0.01</td>
<td>0.082</td>
<td>28.3 ± 7.4</td>
<td>43.8 ± 7.0</td>
<td>11.3 ± 10.8</td>
<td>0.67 ± 0.06</td>
<td></td>
</tr>
<tr>
<td>Operator 2 (Semi-auto)</td>
<td>0.009 ± 0.001</td>
<td>0.014 ± 0.010</td>
<td>0.077</td>
<td>25.8 ± 7.4</td>
<td>39.0 ± 7.8</td>
<td>11.4 ± 9.4</td>
<td>0.69 ± 0.06</td>
<td></td>
</tr>
<tr>
<td>Operator 3 (Semi-auto)</td>
<td>0.009 ± 0.001</td>
<td>0.014 ± 0.010</td>
<td>0.079</td>
<td>26.5 ± 7.1</td>
<td>38.1 ± 7.6</td>
<td>13.9 ± 9.1</td>
<td>0.72 ± 0.06</td>
<td></td>
</tr>
<tr>
<td><strong>Energy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operator 1 (Manual)</td>
<td>0.002 ± 0.001</td>
<td>0.004 ± 0.003</td>
<td>0.032*</td>
<td>27.2 ± 4.7</td>
<td>37.8 ± 6.5</td>
<td>15.8 ± 5.4</td>
<td>0.70 ± 0.05</td>
<td></td>
</tr>
<tr>
<td>Operator 1 (Semi-auto)</td>
<td>0.003 ± 0.001</td>
<td>0.007 ± 0.010</td>
<td>0.131</td>
<td>27.3 ± 6.3</td>
<td>45.3 ± 7.8</td>
<td>7.7 ± 9.0</td>
<td>0.77 ± 0.05</td>
<td></td>
</tr>
<tr>
<td>Operator 2 (Semi-auto)</td>
<td>0.003 ± 0.001</td>
<td>0.007 ± 0.010</td>
<td>0.142</td>
<td>30.9 ± 5.4</td>
<td>49.1 ± 7.5</td>
<td>11.0 ± 8.5</td>
<td>0.72 ± 0.05</td>
<td></td>
</tr>
<tr>
<td>Operator 3 (Semi-auto)</td>
<td>0.003 ± 0.001</td>
<td>0.008 ± 0.010</td>
<td>0.138</td>
<td>30.9 ± 5.8</td>
<td>48.1 ± 8.0</td>
<td>12.2 ± 8.9</td>
<td>0.73 ± 0.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Operator 1 (Manual)</td>
<td>Operator 1 (Semi-auto)</td>
<td>Operator 2 (Semi-auto)</td>
<td>Operator 3 (Semi-auto)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>---------------------</td>
<td>------------------------</td>
<td>------------------------</td>
<td>------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Entropy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>9.35 ± 0.51</td>
<td>8.68 ± 0.31</td>
<td>8.64 ± 0.29</td>
<td>8.58 ± 0.29</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8.62 ± 0.91</td>
<td>8.05 ± 0.86</td>
<td>8.02 ± 0.87</td>
<td>7.96 ± 0.86</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.030*</td>
<td>0.032*</td>
<td>0.036*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>27.5 ± 8.5</td>
<td>24.7 ± 8.3</td>
<td>27.8 ± 6.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>29.5 ± 8.1</td>
<td>25.9 ± 7.9</td>
<td>30.6 ± 7.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>25.3 ± 11.5</td>
<td>23.5 ± 11.1</td>
<td>24.7 ± 8.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.64 ± 0.04</td>
<td>0.73 ± 0.06</td>
<td>0.70 ± 0.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Inertia</strong></td>
<td>Operator 1 (Manual)</td>
<td>Operator 1 (Semi-auto)</td>
<td>Operator 2 (Semi-auto)</td>
<td>Operator 3 (Semi-auto)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>56.82 ± 23.84</td>
<td>44.06 ± 16.49</td>
<td>42.70 ± 16.19</td>
<td>42.10 ± 16.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>47.90 ± 46.02</td>
<td>32.65 ± 17.24</td>
<td>32.64 ± 17.44</td>
<td>31.71 ± 16.96</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.563</td>
<td>0.120</td>
<td>0.166</td>
<td>0.146</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>48.5 ± 12.1</td>
<td>34.2 ± 9.9</td>
<td>37.8 ± 11.1</td>
<td>34.0 ± 10.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>33.4 ± 22.7</td>
<td>21.9 ± 10.6</td>
<td>27.8 ± 13.7</td>
<td>21.8 ± 10.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>65.1 ± 12.2</td>
<td>47.5 ± 14.2</td>
<td>48.8 ± 14.7</td>
<td>47.4 ± 15.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.54 ± 0.05</td>
<td>0.64 ± 0.06</td>
<td>0.60 ± 0.05</td>
<td>0.63 ± 0.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Standard Deviation</strong></td>
<td>Operator 1 (Manual)</td>
<td>Operator 1 (Semi-auto)</td>
<td>Operator 2 (Semi-auto)</td>
<td>Operator 3 (Semi-auto)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>240.46 ± 37.10</td>
<td>197.79 ± 15.92</td>
<td>196.73 ± 14.81</td>
<td>191.18 ± 16.69</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>205.48 ± 62.16</td>
<td>171.15 ± 38.56</td>
<td>169.50 ± 37.89</td>
<td>165.59 ± 36.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.115</td>
<td>0.044*</td>
<td>0.036*</td>
<td>0.042</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>36.1 ± 9.1</td>
<td>36.2 ± 5.4</td>
<td>31.0 ± 4.7</td>
<td>33.6 ± 5.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>27.6 ± 10.4</td>
<td>36.5 ± 6.8</td>
<td>31.3 ± 4.6</td>
<td>29.6 ± 6.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>45.5 ± 12.4</td>
<td>35.8 ± 7.9</td>
<td>30.6 ± 8.9</td>
<td>38.0 ± 7.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.59 ± 0.06</td>
<td>0.58 ± 0.04</td>
<td>0.64 ± 0.04</td>
<td>0.62 ± 0.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Feature values are significant between recurrence and injury groups at the 0.05 level.
Table C2: Individual texture feature values and classification metrics including error, false positive rate (FPR), false negative rate (FNR), and area under the receiver operating characteristic curve (AUC) at 5–8 months post-SABR. All values are reported as the mean ± the standard deviation.

<table>
<thead>
<tr>
<th></th>
<th>Observer</th>
<th>Feature Value</th>
<th>Feature Value</th>
<th>p-value</th>
<th>Error</th>
<th>FPR</th>
<th>FNR</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Recurrence</td>
<td>Injury</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correlation</td>
<td>Operator 1 (Manual)</td>
<td>0.006 ± 0.001</td>
<td>0.009 ± 0.004</td>
<td>0.030*</td>
<td>31.9 ± 5.5</td>
<td>22.7 ± 8.8</td>
<td>43.9 ± 8.1</td>
<td>0.67 ± 0.05</td>
</tr>
<tr>
<td></td>
<td>Operator 1 (Semi-auto)</td>
<td>0.008 ± 0.001</td>
<td>0.012 ± 0.009</td>
<td>0.086</td>
<td>34.3 ± 5.8</td>
<td>26.1 ± 6.0</td>
<td>44.8 ± 10.1</td>
<td>0.65 ± 0.06</td>
</tr>
<tr>
<td></td>
<td>Operator 2 (Semi-auto)</td>
<td>0.008 ± 0.001</td>
<td>0.012 ± 0.009</td>
<td>0.110</td>
<td>41.3 ± 5.6</td>
<td>31.1 ± 6.9</td>
<td>54.7 ± 10.4</td>
<td>0.60 ± 0.06</td>
</tr>
<tr>
<td></td>
<td>Operator 3 (Semi-auto)</td>
<td>0.008 ± 0.001</td>
<td>0.013 ± 0.009</td>
<td>0.086</td>
<td>33.4 ± 5.2</td>
<td>27.4 ± 5.8</td>
<td>41.3 ± 9.6</td>
<td>0.66 ± 0.06</td>
</tr>
<tr>
<td>Energy</td>
<td>Operator 1 (Manual)</td>
<td>0.002 ± 0.001</td>
<td>0.004 ± 0.003</td>
<td>0.014*</td>
<td>27.9 ± 4.8</td>
<td>32.2 ± 6.2</td>
<td>22.4 ± 6.8</td>
<td>0.68 ± 0.04</td>
</tr>
<tr>
<td></td>
<td>Operator 1 (Semi-auto)</td>
<td>0.003 ± 0.001</td>
<td>0.007 ± 0.012</td>
<td>0.192</td>
<td>31.3 ± 5.0</td>
<td>10.2 ± 6.1</td>
<td>58.7 ± 10.3</td>
<td>0.70 ± 0.06</td>
</tr>
<tr>
<td></td>
<td>Operator 2 (Semi-auto)</td>
<td>0.003 ± 0.001</td>
<td>0.007 ± 0.012</td>
<td>0.212</td>
<td>36.5 ± 5.3</td>
<td>11.6 ± 6.7</td>
<td>68.8 ± 12.7</td>
<td>0.66 ± 0.06</td>
</tr>
<tr>
<td></td>
<td>Operator 3 (Semi-auto)</td>
<td>0.003 ± 0.001</td>
<td>0.008 ± 0.012</td>
<td>0.205</td>
<td>35.2 ± 5.1</td>
<td>11.7 ± 6.2</td>
<td>65.8 ± 13.1</td>
<td>0.68 ± 0.07</td>
</tr>
<tr>
<td>Entropy</td>
<td>Operator 1 (Manual)</td>
<td>9.29 ± 0.46</td>
<td>8.48 ± 0.76</td>
<td>0.005*</td>
<td>30.0 ± 4.5</td>
<td>25.2 ± 5.4</td>
<td>36.2 ± 7.4</td>
<td>0.67 ± 0.04</td>
</tr>
<tr>
<td>Operator</td>
<td>Operator Type</td>
<td>Feature 1</td>
<td>Feature 2</td>
<td>Feature 3</td>
<td>Feature 4</td>
<td>Feature 5</td>
<td>Feature 6</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>---------------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td>Operator 1</td>
<td>Semi-auto</td>
<td>8.87 ± 0.44</td>
<td>8.13 ± 0.89</td>
<td>0.019*</td>
<td>22.8 ± 4.8</td>
<td>22.8 ± 2.7</td>
<td>22.8 ± 10.8</td>
<td>0.71 ± 0.05</td>
</tr>
<tr>
<td>Operator 2</td>
<td>Semi-auto</td>
<td>8.73 ± 0.37</td>
<td>8.15 ± 0.93</td>
<td>0.057</td>
<td>31.3 ± 5.0</td>
<td>21.6 ± 4.6</td>
<td>43.9 ± 10.8</td>
<td>0.67 ± 0.07</td>
</tr>
<tr>
<td>Operator 3</td>
<td>Semi-auto</td>
<td>8.68 ± 0.32</td>
<td>8.1 ± 0.92</td>
<td>0.050</td>
<td>28.3 ± 5.7</td>
<td>23.3 ± 3.0</td>
<td>34.7 ± 12.3</td>
<td>0.69 ± 0.07</td>
</tr>
<tr>
<td>Operator 1</td>
<td>Manual</td>
<td>55.14 ± 19.08</td>
<td>52.35 ± 53.76</td>
<td>0.864</td>
<td>43.5 ± 6.4</td>
<td>16.7 ± 9.2</td>
<td>78.4 ± 9.8</td>
<td>0.48 ± 0.06</td>
</tr>
<tr>
<td>Operator 1</td>
<td>Semi-auto</td>
<td>47.77 ± 15.93</td>
<td>34.19 ± 12.88</td>
<td>0.042*</td>
<td>32.8 ± 5.2</td>
<td>16.6 ± 6.1</td>
<td>53.8 ± 10.0</td>
<td>0.60 ± 0.05</td>
</tr>
<tr>
<td>Operator 2</td>
<td>Semi-auto</td>
<td>46.15 ± 15.27</td>
<td>34.46 ± 13.33</td>
<td>0.070</td>
<td>34.9 ± 6.9</td>
<td>20.8 ± 8.0</td>
<td>53.3 ± 12.3</td>
<td>0.58 ± 0.05</td>
</tr>
<tr>
<td>Operator 3</td>
<td>Semi-auto</td>
<td>45.45 ± 13.79</td>
<td>33.52 ± 13.05</td>
<td>0.049*</td>
<td>32.3 ± 7.6</td>
<td>21.1 ± 6.8</td>
<td>46.9 ± 13.5</td>
<td>0.61 ± 0.05</td>
</tr>
<tr>
<td>Operator 1</td>
<td>Manual</td>
<td>240.75 ± 22.24</td>
<td>206.39 ± 39.28</td>
<td>0.016*</td>
<td>33.3 ± 4.8</td>
<td>22.3 ± 7.4</td>
<td>47.5 ± 8.7</td>
<td>0.61 ± 0.04</td>
</tr>
<tr>
<td>Operator 1</td>
<td>Semi-auto</td>
<td>206.57 ± 24.53</td>
<td>175.84 ± 35.26</td>
<td>0.023*</td>
<td>36.8 ± 4.9</td>
<td>24.9 ± 6.8</td>
<td>52.2 ± 9.1</td>
<td>0.61 ± 0.05</td>
</tr>
<tr>
<td>Operator 2</td>
<td>Semi-auto</td>
<td>200.56 ± 17.47</td>
<td>176.56 ± 36.26</td>
<td>0.051</td>
<td>44.0 ± 5.7</td>
<td>31.8 ± 6.8</td>
<td>59.9 ± 11.4</td>
<td>0.55 ± 0.05</td>
</tr>
<tr>
<td>Operator 3</td>
<td>Semi-auto</td>
<td>198.40 ± 16.76</td>
<td>171.94 ± 33.53</td>
<td>0.023*</td>
<td>37.3 ± 4.4</td>
<td>30.0 ± 5.2</td>
<td>46.7 ± 7.9</td>
<td>0.60 ± 0.04</td>
</tr>
</tbody>
</table>

* Feature values are significant between recurrence and injury groups at the 0.05 level.”
APPENDIX D – Curriculum Vitae

EDUCATION

Western University, London, ON
Doctor of Philosophy - Medical Biophysics
(Reclassified from the Master of Science program in Aug. 2012)
Sept. 2011 – present

University of Toronto/The Michener Institute for Applied Health Sciences, Toronto, ON
Bachelor of Science (with Honours) – Medical Radiation Sciences
Diploma in Radiation Therapy
Sept. 2008 – Apr. 2011

Western University, London, ON
Honours Bachelor of Medical Sciences
  • Double major in Medical Sciences and Microbiology & Immunology
Sept. 2004 – Apr. 2008

RESEARCH

Prediction of tumour recurrence after stereotactic ablative radiotherapy (SABR) for lung cancer
MSc/PhD project, Western University
Supervisors: Dr. Aaron Ward & Dr. David Palma
Determine differences in quantitative measures of post-SABR radiological changes on CT between patients with recurrence and radiation fibrosis
Image analysis and machine learning using MATLAB
Feature extraction using Insight Segmentation and Registration Toolkit (ITK) in C++
Statistical analyses in SPSS and MATLAB
Observer study design and implementation for evaluating decision support systems
Sept. 2011 – present

An interprofessional approach to establishing image guided radiotherapy (IGRT) registration guidelines in gynaecological sites
Radiation therapy research methods course
Supervisors: Ms. Caitlin Gillan & Dr. Michael Milosevic
Princess Margaret Hospital, Toronto, ON
Determine inter and intraobserver variability in manual soft tissue image matching in the gynecological site group
Determine the reproducibility and consistency of patient specific matching strategies in IGRT practice
2010 – 2011

Quantifying tumour microvasculature
Medical Biophysics 6-week research project, Western University
Supervisor: Dr. Ian MacDonald
Assisted in analyzing tumour microvasculature images
 Compared quantitative results using statistical analysis
First-hand experience with image analysis through MATLAB
Jan. – Apr. 2006

210
PUBLICATIONS

Refereed journal articles


Letters to the editor

Mattonen, S., Ward, A.D., & Palma, D. In reply: “Raising the bar from target localization toward target characterization: introducing the RTV beyond the GTV.” International Journal of Radiation Oncology, Biology, Physics. (Submitted Mar. 24, 2016)

Refereed conference proceedings (*Indicates presenting author)


Refereed abstracts (*Indicates presenting author)


Mattonen, S.*, Palma, D., Haasbeek, C.J.A., Senan, S., & Ward, A.D. Early prediction of lung cancer recurrence after stereotactic radiotherapy using second-order texture statistics. Canadian Student Health Research Forum. Winnipeg, MB. Jun. 10-12, 2014. (Selected to attend and present a poster by Dr. Andy Watson, the Associate Dean of Graduate and Postdoctoral Studies as one of the top ranked PhD students in the Schulich School of Medicine & Dentistry, received an award of excellence, gold category)


Mattonen, S.*, Gillan, C., Li, W., & Milosevic, M. Communicating clinical information in IGRT practice to facilitate an individualized approach to online image matching guidelines in gynecological sites. RTi3: Radiation Therapy Conference, Toronto, ON. March 4-5, 2011. *(podium presentation)*

**Invited presentations**


Mattonen, S. Prediction of cancer recurrence after stereotactic ablative radiotherapy (SABR) for lung cancer using CT image feature analysis. Department of Medical Biophysics Annual A.C. Burton Day. London, ON. Mar. 21, 2013. *(one of two invited graduate speakers)*
PATENTS


HONOURS AND AWARDS

Natural Sciences and Engineering Council of Canada (NSERC) [$90,000] 2016 – 2018
Postdoctoral Fellowship
$45,000/year for 2 years
Ranked 2/55 in the Computing Sciences Committee

Natural Sciences and Engineering Council of Canada (NSERC) [$105,000] 2013 – 2016
Alexander Graham Bell Canada Graduate Scholarship – Doctoral
$35,000/year for 3 years
Ranked 14/121 in the Computing Sciences Committee

American Society for Radiation Oncology (ASTRO) Annual Meeting Abstract Award [$1,000] 2015
Basic/Translational Science - Junior Investigator Radiation Physics

Dr. Alfred Jay Medical Biophysics Award for Translational Research [$2000] 2016

Imaging Network Ontario Symposium (ImNO) Magna cum Laude Abstract Award 2016
Top rated abstract - Imaging Translation Program

London Health Research Day Top Abstract by Category [$100] 2016
Medical Physics, Engineering and Imaging

Centre for Translational Cancer Research Trainee Travel Award [$1000] 2016
Schulich School of Medicine & Dentistry, Western University

Canadian Cancer Society Research Institute (CCSRI) Travel Award [$1850] 2016

International Conference on the use of Computers in Radiation Therapy (ICCR) - Young Investigator prize shortlist 2016

SPIE Medical Imaging Robert F. Wagner All-Conference Best Student Paper Conference Finalist 2016
Computer-Aided Diagnosis Conference


Canadian Cancer Society Research Institute (CCSRI) Travel Award [$2000] 2015

Nellie Farthing Fellowship, Western University [$3,000] 2015
Recognizing excellence in research to a full-time doctoral student in the Medical Sciences at Western University

American Society for Radiation Oncology (ASTRO) Annual Meeting Abstract Award [$1,000] 2015
Annual Meeting Travel Award in the Physics category

London Imaging Discovery 2nd Place Poster Award [$150] 2014

216
### Canadian Institute of Health Research (CIHR) National Research Poster Competition Award [$500]
- Award of Excellence (Gold category) - Canadian Student Health Research Forum
- 1 of 12 Gold category winners out of approx. 120

### Canadian Institute of Health Research (CIHR) Travel Award [$833.33]
- Canadian Student Health Research Forum - CIHR Poster Presentation

### Imaging Network Ontario Symposium (ImNO) Poster Award [$200]
- 2nd Place Poster - Imaging Translation Program

### 2013-2014 National Volunteer Award Nomination, Let’s Talk Science Outreach

### London Health Research Day Platform Award [$600]
- 2nd place for podium presentation in platform competition – Theatre

### Ontario Graduate Scholarship (OGS) [$15,000]
- Declined

### The Canadian Organization of Medical Physicists (COMP) Young Investigators Award Finalist [$250]

### Oncology Research and Education Day Poster Award [$100]

### Let’s Talk Science – Volunteer of the Month (March)

### Imaging Network Ontario Symposium (ImNO) Poster Award [$200]
- 2nd Place Poster - Imaging Translation Program

### Natural Sciences and Engineering Council of Canada (NSERC) [$17,500]
- Alexander Graham Bell Canada Graduate Scholarship – Master’s

### Ontario Graduate Scholarship (OGS) [$15,000]
- Declined

### London Health Research Day Platform Award [$700]
- 1st place for podium presentation in platform competition – Salon B

### University of Toronto/
The Michener Institute for Applied Health Sciences, Toronto, ON
- Graduated with Honours
  - Faculty Nominated: Excellence in Clinical Preparation Scholarship
    - For consistently exhibiting professional behavior and the ability to work effectively both as an individual and in an inter-professional team in the clinical preparation semester
  - Graduation Award: Outstanding Achievement in Research Methods [$100]
    - Excellence in research by a final year Medical Radiation Sciences student

### Western University, London, ON
- Dean’s Honours List
- Dean’s Honours List
- Western Scholarship of Excellence [$2,000]
EMPLOYMENT HISTORY

Graduate Research Assistant, London, ON
MSc/PhD Student
Supervisors: Dr. Aaron Ward & Dr. David Palma
2011 – present

Let’s Talk Science Outreach Senior Coordinator, London, ON
Western University
Jul. 2015 – present

Teaching Assistant: Medical Biophysics 9520B – Practical Medical Imaging
Supervising weekly labs using the desktop imaging systems
Marking lab reports
Jan – Apr. 2015

Let’s Talk Science Teacher Partnership Coordinator, London, ON
Western University

Ontario Women’s Hockey Association (OWHA) Level II Official, Aurora, ON
Central York Girls Hockey Association
2001 – 2004
2008 – 2010

Wright Dental (WD) Canada, Richmond Hill, ON
Summer Student
2007 – 2008

Vertex Outsourcing, Markham, ON
Summer Student, Human Resources/Application Support
2005 – 2006

SPECIAL TRAINING

Graduate Recruitment Committee
Discuss and prepare/plan for recruitment of new graduate trainees
Participate in graduate student recruitment fairs
Departmental contact with potential graduate students
2014 – 2015

Design of Medical Imaging Labs - Medical Biophysics Course
Gained experience on principles of good teaching and curriculum design
Designed 10 hours of lab material for the DeskCAT optical CT system
These materials were used for course “Medical Biophysics 9520B – Practical Medical Imaging” beginning in Winter 2015
2014

VOLUNTEER EXPERIENCE

Let’s Talk Science, Western University
Organized and demonstrated hands-on science activities, and lead children (preschool through teen) through their completion
Partnered with a science class at Robarts School for the Deaf
Planning committee co-chair for Let’s Talk Cancer high school symposium
Medical Biophysics department representative
Nomination for National Volunteer Award, Let’s Talk Science Outreach
2011 – present
Canadian Cancer Society 2015 – 2016
- Sold daffodil pins during Daffodil Month

Thames Valley Science and Engineering Fair Judge 2015
- Judged grade 6-8 students science fair projects in the physical sciences category
- Team lead for final judging

Community Outreach – Homecoming Medical Biophysics Demonstration 2014
- Provided an explanation and demonstration of the DeskCAT system
- Provided an overview of my thesis to Schulich Medicine alumni

Community Outreach – London Regional Cancer Program Lab Tour 2014
- Provided an explanation and demonstration of the DeskCAT system
- Provided a tour of the Baines Imaging Research Lab

March Break Open House, Western University 2014
- Provided an explanation and demonstration of the DeskCAT system

Consult the Experts, Western University 2013
- Provided support and feedback to graduate/undergraduate students on their research proposals for national and provincial scholarship competitions

- Organized team registration, scheduling, and fundraising for the tournament
- Communicated information regarding schedules, registration, and provided directions to players and families

Peer Tutor, The Michener Institute for Applied Health Sciences 2009 – 2011
- Communicated fundamental and practical concepts of physics and treatment planning to radiation therapy students (three students on a regular basis throughout the two-year period, plus several others on a one-time basis)

Doors Open Toronto, The Michener Institute for Applied Health Sciences 2009
- Provided approx. 25 tours of the computed tomography suites during this one-day event to the general public, and communicated information regarding the equipment and education of technologists using an approach suitable for a lay audience

REVIEWER
- International Journal of Radiation Oncology • Biology • Physics (IJROBP) 2016
- Journal of Medical Imaging 2016

CERTIFICATION
- Canadian Association of Medical Radiation Technologists Jun. 2011
  Registered Radiation Therapist (RTT)