

Imaging texture analysis for automated prediction of lung cancer recurrence after stereotactic radiotherapy

Sarah A. Mattonen
Shyama Tetar
David A. Palma
Alexander V. Louie
Suresh Senan
Aaron D. Ward

Imaging texture analysis for automated prediction of lung cancer recurrence after stereotactic radiotherapy

Sarah A. Mattonen,^{a,*} Shyama Tetar,^b David A. Palma,^{a,c} Alexander V. Louie,^c Suresh Senan,^b and Aaron D. Ward^a

^aThe University of Western Ontario, Department of Medical Biophysics, 1151 Richmond Street, London, Ontario N6A 3K7, Canada

^bVU University Medical Center, Department of Radiation Oncology, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands

^cLondon Regional Cancer Program, Division of Radiation Oncology, 1151 Richmond Street, London, Ontario N6A 3K7, Canada

Abstract. Benign radiation-induced lung injury (RILI) is not uncommon following stereotactic ablative radiotherapy (SABR) for lung cancer and can be difficult to differentiate from tumor recurrence on follow-up imaging. We previously showed the ability of computed tomography (CT) texture analysis to predict recurrence. The aim of this study was to evaluate and compare the accuracy of recurrence prediction using manual region-of-interest segmentation to that of a semiautomatic approach. We analyzed 22 patients treated for 24 lesions (11 recurrences, 13 RILI). Consolidative and ground-glass opacity (GGO) regions were manually delineated. The longest axial diameter of the consolidative region on each post-SABR CT image was measured. This line segment is routinely obtained as part of the clinical imaging workflow and was used as input to automatically delineate the consolidative region and subsequently derive a periconsolidative region to sample GGO tissue. Texture features were calculated, and at two to five months post-SABR, the entropy texture measure within the semiautomatic segmentations showed prediction accuracies [areas under the receiver operating characteristic curve (AUC): 0.70 to 0.73] similar to those of manual GGO segmentations (AUC: 0.64). After integration into the clinical workflow, this decision support system has the potential to support earlier salvage for patients with recurrence and fewer investigations of benign RILI. © 2015 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.JMI.2.4.041010]

Keywords: lung cancer; computed tomography; stereotactic radiation therapy; recurrence; texture analysis; radiomics.

Paper 15072SSPRR received Apr. 1, 2015; accepted for publication Oct. 6, 2015; published online Nov. 12, 2015.

1 Introduction

Stereotactic ablative radiotherapy (SABR) (also known as stereotactic body radiotherapy) is now the guideline-recommended treatment for patients with nonsmall 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 tumors with higher doses per fraction in fewer fractions. Typically, doses of up to 18 Gy per fraction are delivered in between three and eight fractions over one to two weeks, in contrast to a dose of 2 Gy per day delivered over four to six weeks in conventional radiotherapy techniques. The higher doses used in SABR have led to local control rates of up to 90% at three 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 a size and shape similar to those of a recurring tumor 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 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 one 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 tumors. 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, and PET for differentiation of tumor types and 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

*Address all correspondence to: Sarah A. Mattonen, E-mail: smattone@uwo.ca

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 postradiotherapy 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 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 gray-level co-occurrence matrices (GLCM) calculated within manually delineated ground-glass opacity (GGO) regions can predict recurrence within six 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 twofold cross-validation (CV) errors of 23 to 30% and areas under the receiver operating characteristic curve (AUCs) of 0.78 to 0.81.¹³ As seen in Fig. 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 interoperator 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 interoperator variability. Second, manual three-dimensional (3-D) segmentation of the GGO is time-consuming, and automated GGO segmentation is extremely challenging due to the lack of any shape regularity and the difficulty, even for the human medical expert, in judging the locations of the weak GGO boundaries. Inspired by our previous observation that a periconsolidative region (defined by a concentric expansion of the consolidative mass) intended to sample GGO tissue yielded classification performance comparable to manually delineated GGO,¹⁴ we conjectured that accurate classification performance could be obtained using an automatically defined periconsolidative 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 three-month follow-up scan, with a periconsolidative region derived from a semiautomatic 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 semiautomatic, its only input is the Response Evaluation Criteria in Solid Tumors (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 segmentation recurrence predictions to inputs from different operators. Finally, we aim to compare the classification performance to manually delineated segmentations.

2 Methods

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, The 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 the greatest potential value in the clinical context. The proportion of recurrences is artificially inflated in this dataset, and the true rate of local recurrence for a typical stage I NSCLC cohort is ~10%.^{3,4} During follow-up, 46 post-treatment diagnostic CT scans were taken at the first two follow-up time periods (two to five and five to eight months) post-SABR on one of the three scanners at the VU Medical Center: Siemens Volume Zoom 4-slice, Siemens Sensations 64-slice (Siemens Nederland N.V., Den Haag, The Netherlands), or Philips Brilliance iCT 256-slice (Royal Philips Electronics, Inc., Amsterdam, The 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 to 5 mm slice thickness, with 70 cc of intravenous contrast (Ultravist-300; Bayer Pharma AG, Berlin, Germany) administered with a 25 s delay. Iterative reconstruction was not used in this imaging dataset and a standard sharp lung convolution kernel was applied (B60f) with 0.5 to 0.9 mm isotropic in-plane voxel dimensions.

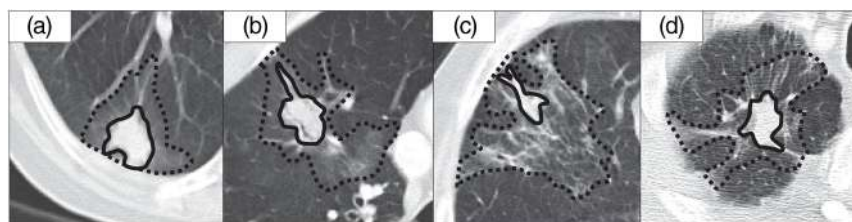


Fig. 1 The two- to five-month follow-up image for four patients in our study. The solid lines encompass consolidative changes and the dashed lines encompass regions of ground-glass opacity (GGO), as delineated by a senior radiation oncology resident. The two lesions that developed benign radiation-induced lung injury are shown in (a) and (b), and qualitatively, the GGO regions have a smooth appearance. (c) and (d) Two lesions that eventually developed cancer recurrence, and in these images, a variegated texture is visible in the GGO regions.

2.2 Region of Interest Segmentation

2.2.1 Manual segmentation of consolidation and ground-glass opacity

Segmentations were performed by a resident in radiation oncology (operator 1) on all follow-up CT images in ITK-SNAP 3.0.¹⁸ 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 tumors 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.

2.2.2 Semiautomatic segmentation of consolidation

An overview of our semiautomatic segmentation approach can be seen in Fig. 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 (ROIs) 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 cut algorithm was used to segment the consolidative regions.^{19,20} 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

$$E(S) = -\beta \|\theta^S - \theta^{\bar{S}}\|_{L1} + |\partial S|, \quad (1)$$

where S is the segmentation, θ^S and $\theta^{\bar{S}}$ are the distributions of object and background intensities, respectively, and β 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, θ^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 four images (two RILI and two recurrences) spanning differences in size, shape, and appearance of the consolidative regions. The input RECIST line segments used to tune the parameter were 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 method²¹ for bin counts >20, allowing for segmentation of objects having more subtle intensity differences from the background.¹⁹ 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 (as described in the Appendix), 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 an ROI centered on and enveloping the background seeds. Any parts of these segmentations lying outside of the whole lung (in

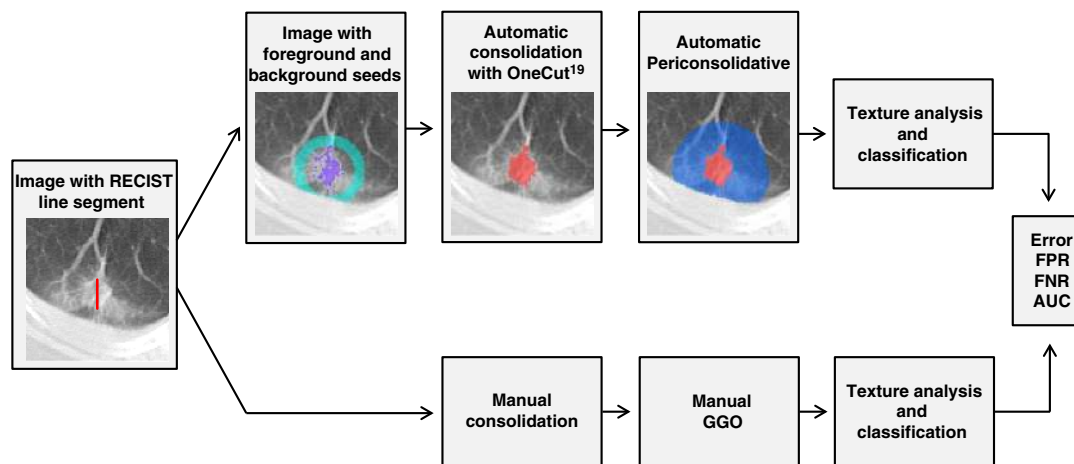


Fig. 2 An overview of our semiautomatic segmentation approach and the methods used in this paper. On the original computed tomography (CT) image, an operator places a Response Evaluation Criteria in Solid Tumors line segment to measure the longest axial diameter (taken in the normal clinical workflow). This was used as initialization for the OneCut algorithm to automatically obtain the consolidation segmentation (red). An expansion of this region defined the periconsolidative region (blue). On the original CT image, manually delineated consolidation and GGO regions were also obtained. Texture analysis was performed in the GGO and semiautomatic periconsolidative regions, and classification results were compared using twofold cross-validation errors, false positive rates (FPR), false negative rates (FNR), and area under the receiver operating characteristic curve (AUC).

cases where the consolidation abutted the lung boundary) were removed and the 3-D largest connected component closest to the RECIST line was taken. Due to the inclusion of small adjacent vessels in the 3-D volume, a subsequent slice-by-slice two-dimensional (2-D) largest connected component was taken to remove these disconnected vessels in plane. Finally, the 3-D largest connected component was taken as the final consolidation segmentation volume.

To validate our algorithm and assess the impact of interoperator 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 semiautomated segmentation algorithm.

2.2.3 Automatic delineation of periconsolidative region

A periconsolidative region within the lung parenchyma was derived automatically by thresholding the 3-D 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 \times 2). This is due to the possible inclusion of connected vessels distant to the consolidative mass or extra consolidative regions erroneously included in the semiautomatic 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 periconsolidative 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 periconsolidative region can be seen in Fig. 2.

2.3 Feature Extraction and Image Analysis

MATLAB[®] 8.4 (The Mathworks, Natick, Massachusetts) was used to calculate first-order texture as the standard deviation of the density within the GGO or periconsolidative region. The Insight Segmentation and Registration Toolkit (ITK) 4.3.1²² was used to calculate four second-order texture features based on a GLCM: energy, entropy, inertia, and correlation.^{23–25} The equations for calculation of these texture features are

$$\text{Energy} = \sum_{i,j} g(i,j)^2, \quad (2)$$

$$\text{Entropy} = -\sum_{i,j} g(i,j) \log_2 g(i,j) \text{ or } 0 \text{ if } g(i,j) = 0, \quad (3)$$

$$\text{Correlation} = \sum_{i,j} \frac{(i - \mu)(j - \mu)g(i,j)}{\sigma^2}, \quad (4)$$

and

$$\text{Inertia} = \sum_{i,j} (i - j)^2 g(i,j), \quad (5)$$

where g is a 2-D matrix where each element $g(i,j)$ contains the number of voxel pairs whose elements have gray levels i and j , where μ is the weighted pixel average,

$$\begin{aligned} \mu &= \sum_{i,j} i \cdot g(i,j) \\ &= \sum_{i,j} j \cdot g(i,j) \text{ (due to symmetry of } g), \end{aligned} \quad (6)$$

and σ is the weighted pixel variance,

$$\begin{aligned} \sigma &= \sum_{i,j} (i - \mu)^2 \cdot g(i,j) \\ &= \sum_{i,j} (j - \mu)^2 \cdot g(i,j) \text{ (due to symmetry of } g). \end{aligned} \quad (7)$$

The number of bins and density ranges in the GLCM were set to yield 20 HU bin widths between -1000 and 200 HU (60 bins), based on analysis of the histogram distributions within manually delineated GGO regions. GLCMs were calculated within four in-plane neighboring voxel pair directions $[(-1,0,0), (-1,-1,0), (0,-1,0), (1,-1,0)]$ for the entire 3-D ROI, and texture features were averaged over all directions. When calculating a GLCM, the neighboring 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 neighboring voxel relationships is shown in Fig. 3. Due to the voxel anisotropy typically seen in postradiation 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: two to five and five to eight months post-SABR. This timing of images was chosen because 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 tumors available.

2.4 Classification

PRTTools 5.0 (Delft Pattern Recognition Research, Delft, The Netherlands) was used for classification. For a stringent

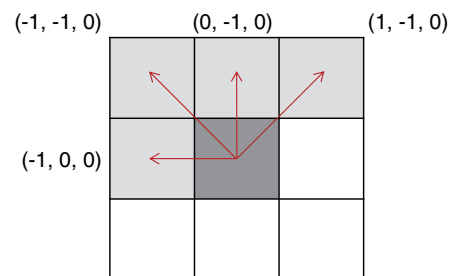


Fig. 3 The four in-plane spatial relationships used for calculating the gray level co-occurrence matrix. The dark central voxel is the reference voxel, and the neighboring voxels considered in the analysis are shown in gray.

determination of classification performance, twofold CV over 1000 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 periconsolidative regions and the manually delineated GGO regions.

2.5 Comparison of the Semiautomatic Segmentations

To compare the segmented regions between operators, similarity metrics were calculated to measure segmentation differences. The symmetric mean absolute boundary difference, Dice similarity coefficient (DSC) $[(2|V_A \cap V_B|)/(|V_A| + |V_B|)]$, volume difference $(V_B - V_A)$, recall $[TP/(TP + FN)]$, and precision $[TP/(TP + FP)]$ were calculated, where V_A and V_B are the volumes of the segmentations by operators A and B, respectively, and TP = true positive, FP = false positive, TN = true negative, and FN = false negative. Segmentation metrics were calculated on all of the images used in this study. Each metric was calculated on the same 3-D image volume between operators. Calculation of segmentation differences was completed between all operators for the semiautomated 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 $\{2 \cdot [(precision \cdot recall)/(precision + recall)]\}$. Additionally, the time to generate the manual and semiautomatic segmentations was measured and compared.

2.6 Statistical Analysis

All statistical analyses were completed in SPSS Statistics Version 22.0 (IBM Corp., Armonk, New York). 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 $\alpha \leq 0.05$). 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 CV metrics (1000 times), to correct for multiple testing, a Bonferroni correction was applied with $\alpha \leq 0.05/1000$. To determine noninferiority of the semiautomated classification results with respect to results obtained using the manual segmentations, the 95% confidence interval of the difference in CV metrics was assessed. An inferiority margin of 5% was chosen as an acceptable clinical difference in CV metrics.

3 Results

3.1 Manual Segmentations

Classification performance at two to five months post-SABR using texture features within manually delineated GGO regions is shown in Fig. 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 twofold CV error was entropy. At two to five months post-SABR, the mean twofold CV error was 27.4%,

Table 1 Area under the receiver operating characteristic curve for the top texture feature (entropy) at both time points after stereotactic ablative radiotherapy (SABR).

	Two to five months	Five to eight months
Operator 1 (manual)	0.64 ± 0.04	0.67 ± 0.04
Operator 1 (semiauto)	0.73 ± 0.06	0.71 ± 0.05
Operator 2 (semiauto)	0.70 ± 0.04	0.67 ± 0.07
Operator 3 (semiauto)	0.71 ± 0.05	0.69 ± 0.07

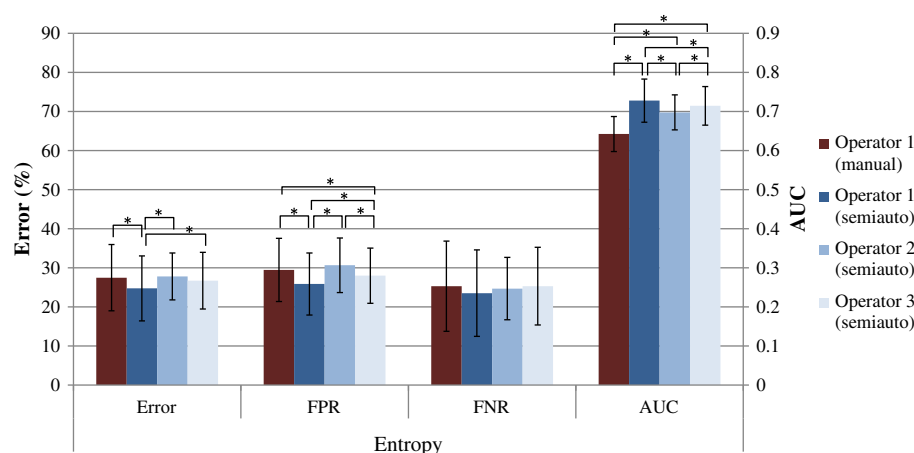


Fig. 4 Classification performance of the manual GGO and semiautomatic periconsolidative regions for the top-performing texture feature (entropy) at two to five months post-SABR. The columns indicate the mean twofold cross-validation errors, mean FPR, mean FNR, and the AUC. The whiskers indicate the standard deviation over 1000 repetitions, and each color represents a different operator. The asterisks indicate a statistically significant difference between operators ($p < 0.00005$).

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 five to eight months post-SABR was 0.67 and is shown in Table 1. At five to eight months post-SABR, the entropy feature demonstrated a twofold CV error of 30.0%, mean FPR of 25.2%, and mean FNR of 36.2%. A qualitative representation of the manually delineated ROIs is provided in Fig. 1.

3.2 Semiautomated Segmentations

3.2.1 Segmentation comparison

Similarity metrics for the semiautomatic consolidative segmentations are summarized in Table 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 semiautomatic approach, interoperator variability was low, with values for the F_1 score ranging from 0.87 to 0.93 and boundary differences from 1.00 to 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 to 0.93. Qualitative

examples for all of the operators' semiautomatic segmentations are shown in Fig. 5.

3.2.2 Classification performance

Classification performance at two to five months post-SABR within the automatically derived periconsolidative regions is shown in Fig. 4 for the top feature (entropy). Using these segmentations, entropy had mean twofold CV errors of 24.7 to 27.8% across all three operators. Operator 1 had a significantly lower mean twofold 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 semiautomatic segmentations produced balanced FPRs and FNRs of 25.8 to 30.6% and 23.5 to 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 $\leq 5\%$. At two to five months post-SABR, entropy had AUC values between 0.70 and 0.73 for all operators, which was significantly different between all operators. For comparison, the AUC values at five to eight months post-SABR are shown in Table 1 and were between 0.67 and 0.71 for all operators. At this time point, the entropy feature had twofold CV errors of 22.8 to 31.3%, mean FPRs of 21.6 to 23.3%, and mean FNRs of 22.8 to 43.9%.

Table 2 Interoperator variability in the semiautomatic consolidative segmentations; all values are reported as the median [interquartile range].

Operator	MAD bilateral (mm)	DSC	F_1 score	Volume difference (cm ³)
1 versus 2	1.18 [0.31, 2.62]	0.90 [0.78, 0.98]	0.90 [0.78, 0.98]	-1.66 [-17.91, 0.33] [#]
1 versus 3	1.75 [0.60, 3.19] [#]	0.87 [0.74, 0.97] [#]	0.87 [0.74, 0.97] [#]	-10.56 [-22.86, -1.68] ^{**}
2 versus 3	1.00 [0.45, 2.13] [#]	0.93 [0.81, 0.97] [#]	0.93 [0.81, 0.97] [#]	-3.57 [-13.58, 0.06] [#]

Note: MAD, mean absolute boundary difference; DSC, Dice similarity coefficient.

*[#]Indicate significant differences between rows, $p < 0.05$.

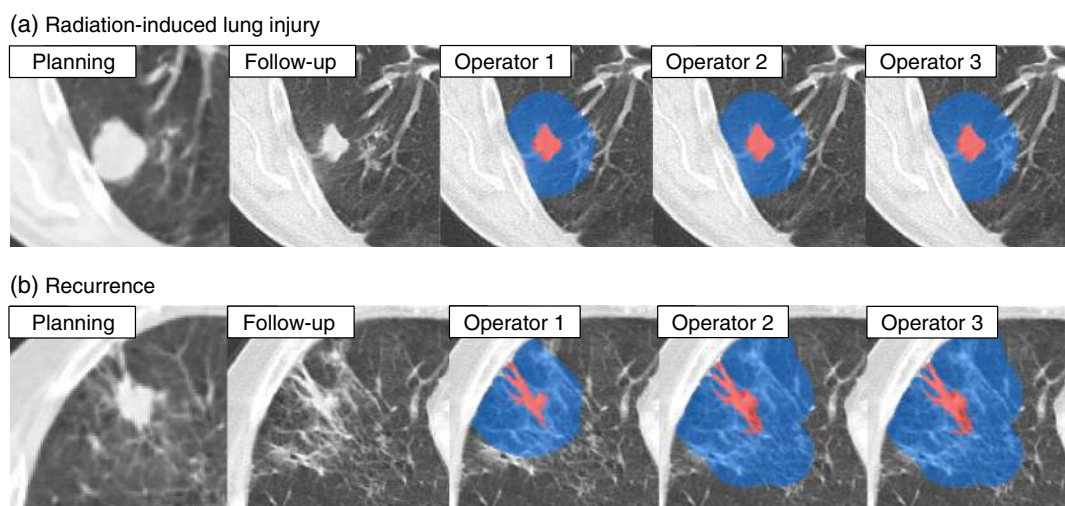


Fig. 5 The pretreatment lesion used for treatment planning and a three-month follow-up scan showing the radiological changes post-SABR for a patient with (a) radiation-induced lung injury and (b) recurrence. Semiautomated segmentations of the consolidative regions (red) and periconsolidative regions (blue) surround them. Operators 1, 2, and 3 are shown for each case.

3.3 Comparison of System Performance: Manual versus Semiautomatic Segmentations

3.3.1 Classification performance

Comparison of classification performance at two to five months post-SABR using the manual and semiautomatic decision support systems is shown in Fig. 4. Classification using the semiautomated method was compared to classification performance using the manual delineations performed by the senior radiation oncology resident. Using the semiautomatic segmentations, the entropy texture feature produced AUCs higher than the manual contours for all three operators. All AUC values using the semiautomated approach were significantly different from reference manual contours. In terms of mean twofold error, operator 1 demonstrated a significantly better performance compared to the manual segmentations. There was no significant difference in classification performance between the manual and the semiautomatic 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 noninferiority study was completed to determine if the semiautomated approach was noninferior to the manual approach. Figure 6 shows the 95% confidence intervals of the differences between the semiautomated and manual classification metrics. Noninferiority was demonstrated for all metrics as the 95% confidence intervals fell within clinically acceptable differences of 5%. At five to eight months post-SABR, the AUC and twofold error values were similar to those at two to five months. Results were also consistent between the manual and semiautomated systems, with AUC values within 0.04 and twofold errors within 7%. Overall, classification results showed robustness using the

semiautomatic 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 Tables 3 and 4.

3.3.2 Timing

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

4 Discussion

The ability to distinguish benign fibrosis from tumor recurrence is crucial in determining a patient's care following SABR and determining 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 previous work has shown that the entropy texture feature calculated within manually delineated regions of GGO could predict recurrence with twofold CV errors of 24% at two to five 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 Fig. 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 the 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 semiautomatic segmentations.

This work has shown the ability to predict recurrence post-SABR using texture analysis in regions delineated by means of a semiautomated 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 pretreatment 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 ROIs, regardless of deformations and retractions from the pretreatment 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

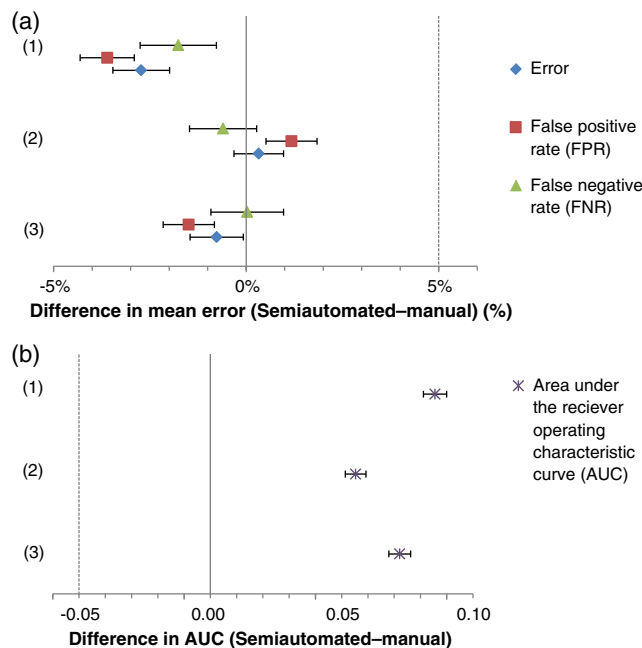


Fig. 6 Noninferiority analysis for the difference between means (semiautomated-manual) for the (a) classification error metrics and (b) AUC. Each numbered row of each figure corresponds to a single operator. The markers indicate the mean differences and the whiskers represent the 95% confidence intervals. The vertical dashed lines represent an acceptable clinical difference of 5% for all metrics.

Table 3 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 two to five months post-SABR. All values are reported as mean \pm standard deviation.

	Observer	Feature value— recurrence	Feature value— injury	<i>p</i> value	Error	FPR	FNR	AUC
Correlation	Operator 1 (manual)	0.006 \pm 0.002	0.010 \pm 0.007	0.098	33.6 \pm 10.9	37.8 \pm 12.9	29.0 \pm 12.8	0.61 \pm 0.06
	Operator 1 (semiauto)	0.008 \pm 0.001	0.014 \pm 0.010	0.082	28.3 \pm 7.4	43.8 \pm 7.0	11.3 \pm 10.8	0.67 \pm 0.06
	Operator 2 (semiauto)	0.009 \pm 0.001	0.014 \pm 0.010	0.077	25.8 \pm 7.4	39.0 \pm 7.8	11.4 \pm 9.4	0.69 \pm 0.06
	Operator 3 (semiauto)	0.009 \pm 0.001	0.014 \pm 0.010	0.079	26.5 \pm 7.1	38.1 \pm 7.6	13.9 \pm 9.1	0.72 \pm 0.06
Energy	Operator 1 (manual)	0.002 \pm 0.001	0.004 \pm 0.003	0.032*	27.2 \pm 4.7	37.8 \pm 6.5	15.8 \pm 5.4	0.70 \pm 0.05
	Operator 1 (semiauto)	0.003 \pm 0.001	0.007 \pm 0.010	0.131	27.3 \pm 6.3	45.3 \pm 7.8	7.7 \pm 9.0	0.77 \pm 0.05
	Operator 2 (semiauto)	0.003 \pm 0.001	0.007 \pm 0.010	0.142	30.9 \pm 5.4	49.1 \pm 7.5	11.0 \pm 8.5	0.72 \pm 0.05
	Operator 3 (semiauto)	0.003 \pm 0.001	0.008 \pm 0.010	0.138	30.9 \pm 5.8	48.1 \pm 8.0	12.2 \pm 8.9	0.73 \pm 0.06
Entropy	Operator 1 (manual)	9.35 \pm 0.51	8.62 \pm 0.91	0.030*	27.5 \pm 8.5	29.5 \pm 8.1	25.3 \pm 11.5	0.64 \pm 0.04
	Operator 1 (semiauto)	8.68 \pm 0.31	8.05 \pm 0.86	0.032*	24.7 \pm 8.3	25.9 \pm 7.9	23.5 \pm 11.1	0.73 \pm 0.06
	Operator 2 (semiauto)	8.64 \pm 0.29	8.02 \pm 0.87	0.036*	27.8 \pm 6.0	30.6 \pm 7.0	24.7 \pm 8.0	0.70 \pm 0.04
	Operator 3 (semiauto)	8.58 \pm 0.29	7.96 \pm 0.86	0.035*	26.7 \pm 7.3	28.0 \pm 7.1	25.3 \pm 9.9	0.71 \pm 0.05
Inertia	Operator 1 (manual)	56.82 \pm 23.84	47.90 \pm 46.02	0.563	48.5 \pm 12.1	33.4 \pm 22.7	65.1 \pm 12.2	0.54 \pm 0.05
	Operator 1 (semiauto)	44.06 \pm 16.49	32.65 \pm 17.24	0.120	34.2 \pm 9.9	21.9 \pm 10.6	47.5 \pm 14.2	0.64 \pm 0.06
	Operator 2 (semiauto)	42.70 \pm 16.19	32.64 \pm 17.44	0.166	37.8 \pm 11.1	27.8 \pm 13.7	48.8 \pm 14.7	0.60 \pm 0.05
	Operator 3 (semiauto)	42.10 \pm 16.04	31.71 \pm 16.96	0.146	34.0 \pm 10.1	21.8 \pm 10.4	47.4 \pm 15.1	0.63 \pm 0.06
Standard deviation	Operator 1 (manual)	240.46 \pm 37.10	205.48 \pm 62.16	0.115	36.1 \pm 9.1	27.6 \pm 10.4	45.5 \pm 12.4	0.59 \pm 0.06
	Operator 1 (semiauto)	197.79 \pm 15.92	171.15 \pm 38.56	0.044*	36.2 \pm 5.4	36.5 \pm 6.8	35.8 \pm 7.9	0.58 \pm 0.04
	Operator 2 (semiauto)	196.73 \pm 14.81	169.50 \pm 37.89	0.036*	31.0 \pm 4.7	31.3 \pm 4.6	30.6 \pm 8.9	0.64 \pm 0.04
	Operator 3 (semiauto)	191.18 \pm 16.69	165.59 \pm 36.13	0.042	33.6 \pm 5.6	29.6 \pm 6.8	38.0 \pm 7.9	0.62 \pm 0.04

Note: *Feature values are significant between recurrence and injury groups at the 0.05 level.

words, we would obtain this line segment when we obtain the follow-up image, leading to a fully automatic overall system based on semiautomated segmentation.

We reported small differences between the semiautomated 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.^{26,27} 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 sizes and shapes of the consolidative masses (they are not typically spherical, as is common in the case of pretreatment tumors) as shown in Fig. 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 cut algorithm, since the foreground and background samples (as shown in

Fig. 2) are likely to vary only slightly. The F_1 scores were all >0.85 , suggesting high interoperator agreement in the semiautomatic 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 nonsmooth and highly variable shapes of the consolidative regions (as seen in Fig. 1). Most importantly, however, this semiautomatic segmentation provides a basis for the automatically defined periconsolidative region. As shown in Fig. 5, the periconsolidative regions are in general dissimilar to the manual GGO regions seen in Fig. 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

Table 4 Individual texture feature values and classification metrics including error, FPR, FNR, and AUC at five to eight months post-SABR. All values are reported as mean \pm standard deviation.

	Observer	Feature value— recurrence	Feature value— injury	<i>p</i> value	Error	FPR	FNR	AUC
Correlation	Operator 1 (manual)	0.006 \pm 0.001	0.009 \pm 0.004	0.030*	31.9 \pm 5.5	22.7 \pm 8.8	43.9 \pm 8.1	0.67 \pm 0.05
	Operator 1 (semiauto)	0.008 \pm 0.001	0.012 \pm 0.009	0.086	34.3 \pm 5.8	26.1 \pm 6.0	44.8 \pm 10.1	0.65 \pm 0.06
	Operator 2 (semiauto)	0.008 \pm 0.001	0.012 \pm 0.009	0.110	41.3 \pm 5.6	31.1 \pm 6.9	54.7 \pm 10.4	0.60 \pm 0.06
	Operator 3 (semiauto)	0.008 \pm 0.001	0.013 \pm 0.009	0.086	33.4 \pm 5.2	27.4 \pm 5.8	41.3 \pm 9.6	0.66 \pm 0.06
Energy	Operator 1 (manual)	0.002 \pm 0.001	0.004 \pm 0.003	0.014*	27.9 \pm 4.8	32.2 \pm 6.2	22.4 \pm 6.8	0.68 \pm 0.04
	Operator 1 (semiauto)	0.003 \pm 0.001	0.007 \pm 0.012	0.192	31.3 \pm 5.0	10.2 \pm 6.1	58.7 \pm 10.3	0.70 \pm 0.06
	Operator 2 (semiauto)	0.003 \pm 0.001	0.007 \pm 0.012	0.212	36.5 \pm 5.3	11.6 \pm 6.7	68.8 \pm 12.7	0.66 \pm 0.06
	Operator 3 (semiauto)	0.003 \pm 0.001	0.008 \pm 0.012	0.205	35.2 \pm 5.1	11.7 \pm 6.2	65.8 \pm 13.1	0.68 \pm 0.07
Entropy	Operator 1 (manual)	9.29 \pm 0.46	8.48 \pm 0.76	0.005*	30.0 \pm 4.5	25.2 \pm 5.4	36.2 \pm 7.4	0.67 \pm 0.04
	Operator 1 (semiauto)	8.87 \pm 0.44	8.13 \pm 0.89	0.019*	22.8 \pm 4.8	22.8 \pm 2.7	22.8 \pm 10.8	0.71 \pm 0.05
	Operator 2 (semiauto)	8.73 \pm 0.37	8.15 \pm 0.93	0.057	31.3 \pm 5.0	21.6 \pm 4.6	43.9 \pm 10.8	0.67 \pm 0.07
	Operator 3 (semiauto)	8.68 \pm 0.32	8.10 \pm 0.92	0.050	28.3 \pm 5.7	23.3 \pm 3.0	34.7 \pm 12.3	0.69 \pm 0.07
Inertia	Operator 1 (manual)	55.14 \pm 19.08	52.35 \pm 53.76	0.864	43.5 \pm 6.4	16.7 \pm 9.2	78.4 \pm 9.8	0.48 \pm 0.06
	Operator 1 (semiauto)	47.77 \pm 15.93	34.19 \pm 12.88	0.042*	32.8 \pm 5.2	16.6 \pm 6.1	53.8 \pm 10.0	0.60 \pm 0.05
	Operator 2 (semiauto)	46.15 \pm 15.27	34.46 \pm 13.33	0.070	34.9 \pm 6.9	20.8 \pm 8.0	53.3 \pm 12.3	0.58 \pm 0.05
	Operator 3 (semiauto)	45.45 \pm 13.79	33.52 \pm 13.05	0.049*	32.3 \pm 7.6	21.1 \pm 6.8	46.9 \pm 13.5	0.61 \pm 0.05
Standard deviation	Operator 1 (manual)	240.75 \pm 22.24	206.39 \pm 39.28	0.016*	33.3 \pm 4.8	22.3 \pm 7.4	47.5 \pm 8.7	0.61 \pm 0.04
	Operator 1 (semiauto)	206.57 \pm 24.53	175.84 \pm 35.26	0.023*	36.8 \pm 4.9	24.9 \pm 6.8	52.2 \pm 9.1	0.61 \pm 0.05
	Operator 2 (semiauto)	200.56 \pm 17.47	176.56 \pm 36.26	0.051	44.0 \pm 5.7	31.8 \pm 6.8	59.9 \pm 11.4	0.55 \pm 0.05
	Operator 3 (semiauto)	198.40 \pm 16.76	171.94 \pm 33.53	0.023*	37.3 \pm 4.4	30.0 \pm 5.2	46.7 \pm 7.9	0.60 \pm 0.04

Note: *Feature values are significant between recurrence and injury groups at the 0.05 level.

Our system shows high predictive accuracy using the second-order entropy texture feature, with AUCs of 0.70 to 0.73 at two to five months post-SABR. It was also robust for interoperator variability in the initialization of the system by placement of the RECIST diameter measurement. Although classification performance using the semiautomated method was compared only against a single operator's manual segmentations, noninferiority 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 semiautomatic 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 five to eight months post-SABR were comparable to the results at two to five months post-SABR. This demonstrates stability of system performance through time, and also that the appearance of radiological

changes at the two earliest clinically scheduled follow-up time points (approximately three and six 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 Fig. 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 after treatment.⁶ Other imaging-based features, such as size or mass-like shape, can be ineffective for early detection of local recurrence post-SABR.^{28–30} FDG-PET has also been investigated for distinguishing fibrosis from recurrence post-SABR, and it has been shown that maximum

standardized uptake values can be predictive of recurrence, but not until a year post-SABR.^{31,32} 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.³³ 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 of semiautomatic 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 tumor 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 dataset. The focus of this study was on the early prediction of recurrence and images analyzed at three and six months post-SABR. Further studies should be completed on additional time points to determine the usefulness of image features for recurrence prediction as post-SABR time 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.³⁴ 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 semiautomated algorithm was completed with only three 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 type of algorithm. Another limitation of the current work is the use of the same imaging dataset as used in our previous publications. 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.

5 Conclusion

Second-order texture features calculated within GGO delineated from a semiautomated 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 six months of SABR. At two to five months post-SABR, second-order entropy provided good recurrence prediction based on semiautomatic segmentations, with AUCs of 0.70 to 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 segmentation 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 could 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.

Appendix

The regions of the image containing normal lung parenchyma can be obtained from the planning computed tomography 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.^{35–38} Although this step is outside of the scope of our method, for the purposes of reproducibility, we are providing in this appendix 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.³⁹ For trachea-adjacent tumors, the trachea was separately segmented superior to the carina. A three-dimensional rectilinear region of interest was defined encompassing the segmentation target (the entire lung volume or the trachea) and preprocessed 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 as previously described.^{14,40} 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. Postprocessing was performed in MATLAB® 8.4 (The Mathworks Inc., Natick, Massachusetts) by a slice-wise morphological filling of small vessels or artifacts <10 mm² 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.

Acknowledgments

The authors would like to acknowledge the following sources of funding: the Natural Sciences and Engineering Research Council of Canada, the Ontario Institute for Cancer Research, and Cancer Care Ontario. A. D. Ward holds a Cancer Care Ontario Research Chair in Cancer Imaging. The authors would also like to acknowledge Maysam Shahedi for his assistance in calculation of segmentation similarity metrics and Carol Johnson for her assistance in software programming.

References

1. D. S. Ettinger et al., "Non-small cell lung cancer, version 1.2015," *J. Natl. Compr. Cancer Netw.* **12**(12), 1738–1761 (2014).
2. J. Vansteenkiste et al., "Early and locally advanced non-small-cell lung cancer (NSCLC): ESMO clinical practice guidelines for diagnosis, treatment and follow-up," *Ann. Oncol.* **24**(suppl 6), vi89–vi98 (2013).
3. N. P. Nguyen et al., "Can stereotactic fractionated radiation therapy become the standard of care for early stage non-small cell lung carcinoma," *Cancer Treat. Rev.* **34**(8), 719–727 (2008).
4. F. J. Lagerwaard et al., "Outcomes of stereotactic ablative radiotherapy in patients with potentially operable stage I non-small cell lung cancer," *Int. J. Radiat. Oncol. Biol. Phys.* **83**(1), 348–353 (2012).
5. J. Vansteenkiste et al., "2nd ESMO Consensus Conference on Lung Cancer: early-stage non-small-cell lung cancer consensus on diagnosis, treatment and follow-up," *Ann. Oncol.* **25**(8), 1462–1474 (2014).
6. K. Huang et al., "High-risk CT features for detection of local recurrence after stereotactic ablative radiotherapy for lung cancer," *Radiother. Oncol.* **109**(1), 51–57 (2013).
7. K. Huang et al., "Radiographic changes after lung stereotactic ablative radiotherapy (SABR)—can we distinguish recurrence from fibrosis? A systematic review of the literature," *Radiother. Oncol.* **102**(3), 335–342 (2012).
8. L. Alic, W. J. Niessen, and J. F. Veenland, "Quantification of heterogeneity as a biomarker in tumor imaging: a systematic review," *PLoS One* **9**(10), e110300 (2014).
9. S. Alobaidli et al., "The role of texture analysis in imaging as an outcome predictor and potential tool in radiotherapy treatment planning," *Br. J. Radiol.* **87**(1042), 20140369 (2014).
10. S. Krafft et al., "TU-CD-BRB-01: normal lung CT texture features improve predictive models for radiation pneumonitis," *Med. Phys.* **42**(6), 3602 (2015).
11. A. R. Cunliffe et al., "Lung texture in serial thoracic CT scans: correlation with radiologist-defined severity of acute changes following radiation therapy," *Phys. Med. Biol.* **59**(18), 5387 (2014).
12. A. Cunliffe et al., "Lung texture in serial thoracic computed tomography scans: correlation of radiomics-based features with radiation therapy dose and radiation pneumonitis development," *Int. J. Radiat. Oncol. Biol. Phys.* **91**(5), 1048–1056 (2015).
13. S. A. Mattonen et al., "Early prediction of tumor recurrence based on CT texture changes after stereotactic ablative radiotherapy (SABR) for lung cancer," *Med. Phys.* **41**(3), 033502 (2014).
14. S. A. Mattonen et al., "Early prediction of lung cancer recurrence after stereotactic radiotherapy using second order texture statistics," *Proc. SPIE* **9038**, 90380T (2014).
15. A. V. Louie et al., "Inter-observer and intra-observer reliability for lung cancer target volume delineation in the 4D-CT era," *Radiother. Oncol.* **95**(2), 166–171 (2010).
16. S. A. Mattonen et al., "Texture analysis of automatic graph cuts segmentations for detection of lung cancer recurrence after stereotactic radiotherapy," *Proc. SPIE* **9417**, 941719 (2015).
17. E. A. Eisenhauer et al., "New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1)," *Eur. J. Cancer* **45**(2), 228–247 (2009).
18. P. A. Yushkevich et al., "User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability," *Neuroimage* **31**(3), 1116–1128 (2006).
19. M. Tang et al., "GrabCut in one cut," in *IEEE Int. Conf. on Computer Vision*, pp. 1769–1776, IEEE (2013).
20. Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Trans. Pattern Anal. Mach. Intell.* **26**(9), 1124–1137 (2004).
21. C. Rother, V. Kolmogorov, and A. Blake, "GrabCut: interactive foreground extraction using iterated graph cuts," *ACM Trans. Graph.* **23**(3), 309–314 (2004).
22. T. S. Yoo et al., "Engineering and algorithm design for an image processing API: a technical report on ITK-the insight toolkit," *Stud. Health Technol. Inform.* **85**, 586–592 (2002).
23. R. W. Connors and C. A. Harlow, "A theoretical comparison of texture algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-2**(3), 204–222 (1980).
24. R. W. Connors, M. M. Trivedi, and C. A. Harlow, "Segmentation of a high-resolution urban scene using texture operators," *Comput. Vis. Graph. Image Process.* **25**(3), 273–310 (1984).
25. R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst. Man Cybern.* **SMC-3**(6), 610–621 (1973).
26. J. J. Erasmus et al., "Interobserver and intraobserver variability in measurement of non-small-cell carcinoma lung lesions: implications for assessment of tumor response," *J. Clin. Oncol.* **21**(13), 2574–2582 (2003).
27. C. Suzuki et al., "Interobserver and intraobserver variability in the response evaluation of cancer therapy according to RECIST and WHO-criteria," *Acta Oncol.* **49**(4), 509–514 (2010).
28. N. E. Dunlap et al., "Computed tomography-based anatomic assessment overestimates local tumor recurrence in patients with mass-like consolidation after stereotactic body radiotherapy for early-stage non-small cell lung cancer," *Int. J. Radiat. Oncol. Biol. Phys.* **84**(5), 1071–1077 (2012).
29. D. B. Shultz et al., "Imaging features associated with disease progression after stereotactic ablative radiotherapy for stage I non-small-cell lung cancer," *Clin. Lung Cancer* **15**(4), 294–301 (2014).
30. A. Takeda et al., "Evaluation for local failure by 18F-FDG PET/CT in comparison with CT findings after stereotactic body radiotherapy (SBRT) for localized non-small-cell lung cancer," *Lung Cancer* **79**(3), 248–253 (2013).
31. N. Nakajima et al., "Differentiation of tumor recurrence from radiation-induced pulmonary fibrosis after stereotactic ablative radiotherapy for lung cancer: characterization of 18F-FDG PET/CT findings," *Ann. Nucl. Med.* **27**(3), 261–270 (2013).
32. M. Essler et al., "Positron-emission tomography CT to identify local recurrence in stage I lung cancer patients 1 year after stereotactic body radiation therapy," *Strahlenther. Oncol.* **189**(6), 495–501 (2013).
33. B. Sawyer et al., "CT perfusion imaging in response assessment of pulmonary metastases undergoing stereotactic ablative radiotherapy," *J. Med. Imaging Radiat. Oncol.* **59**(2), 207–215 (2015).
34. V. Kumar et al., "Radiomics: the process and the challenges," *Magn. Reson. Imaging* **30**(9), 1234–1248 (2012).
35. G. Gill, M. Toews, and R. R. Beichel, "Robust initialization of active shape models for lung segmentation in CT scans: a feature-based atlas approach," *Int. J. Biomed. Imaging* **2014**, 479154 (2014).
36. A. Mansoor et al., "A generic approach to pathological lung segmentation," *IEEE Trans. Med. Imaging* **33**(12), 2293–2310 (2014).
37. M. N. Prasad et al., "Automatic segmentation of lung parenchyma in the presence of diseases based on curvature of ribs," *Acad. Radiol.* **15**(9), 1173–1180 (2008).
38. N. Birkbeck et al., *Lung Segmentation from CT with Severe Pathologies Using Anatomical Constraints*, Springer, New York (2014).
39. S. C. Zhu, T. S. Lee, and A. L. Yuille, "Region competition: unifying snakes, region growing, energy/Bayes/MDL for multi-band image segmentation," in *Proc. Fifth Int. Conf. on Computer Vision*, pp. 416–423, IEEE (1995).
40. R. Whitaker, "A level-set approach to 3D reconstruction from range data," *Int. J. Comput. Vis.* **29**(3), 203–231 (1998).

Sarah A. Mattonen is a PhD candidate in the Department of Medical Biophysics at Western University in London, Canada. She holds an honors BSc from Western University in Medical Sciences and a BSc/diploma in radiation therapy from the University of Toronto. She has received funding from the Natural Sciences and Engineering Research Council of Canada (NSERC) and currently holds an Alexander Graham Bell Doctoral Canada Graduate Scholarship.

Shyama Tatar is a senior resident in the Department of Radiation Oncology at the VU University Medical Center in Amsterdam, The Netherlands. During her residency she has been active in lung cancer research in the field of radiotherapy, with a focus on thoracic high-dose re-irradiation, lung cancer recurrence after stereotactic radiotherapy (SBRT) and synchronous treatment of multiple lung tumors with SBRT. She is a member of ASTRO and the International Association for the Study of Lung Cancer.

David A. Palma is a radiation oncologist at the London Health Sciences Centre and a clinician-scientist with the Ontario Institute for Cancer Research. His research focuses on new imaging technologies used with radiotherapy to improve tumour targeting and deliver higher doses of radiotherapy than previously possible, and new imaging methods that can better assess response after treatment.

Alexander V. Louie is a radiation oncologist and clinician-scientist specializing in thoracic malignancies at the London Regional Cancer Program. He received a MSc degree in epidemiology at Harvard University, where he developed a special interest in health economics and comparative effectiveness research in the context of lung cancer. He will be obtaining his PhD "Stereotactic ablative

radiotherapy for Stage I Non-small Cell Lung Cancer: Optimizing Patient Selection through Comparative Effectiveness research" later this year.

Suresh Senan is a radiation oncologist and professor of clinical experimental radiotherapy at the VU University Medical Center in Amsterdam, The Netherlands. His research group has been active in image-guided and stereotactic radiotherapy, 4-dimensional CT, chemo-radiotherapy, small-cell lung cancer and comparative effectiveness research. He is a member of ASTRO, the International Association for the Study of Lung Cancer, ASCO, ESTRO, and the European Society for Medical Oncology.

Aaron D. Ward is an assistant professor of Medical Biophysics, Biomedical Engineering, and Oncology at Western University in London, Canada. His research focuses on development and clinical translation of computational techniques for diagnosis, therapy guidance, and treatment response assessment in oncology. He has received funding from the Natural Sciences and Engineering Research Council of Canada, the Canadian Institutes of Health Research, the Ontario Institute for Cancer Research, Prostate Cancer Canada, and Cancer Care Ontario.