

PAPER

Association of pre-treatment radiomic features with lung cancer recurrence following stereotactic body radiation therapy

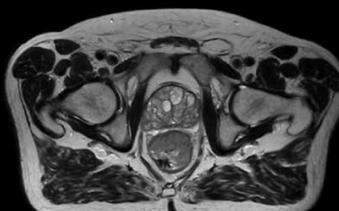
To cite this article: Kyle J Lafata *et al* 2019 *Phys. Med. Biol.* **64** 025007

View the [article online](#) for updates and enhancements.

Uncompromised.

See clearly during treatment to attack the tumor and protect the patient.

Two worlds, one future.



Captured on Elekta high-field MR-linac during 2018 imaging studies.

 **Elekta**

Elekta MR-linac is pending 510(k) premarket clearance and not available for commercial distribution or sale in the U.S.



PAPER

Association of pre-treatment radiomic features with lung cancer recurrence following stereotactic body radiation therapy

RECEIVED
5 November 2018ACCEPTED FOR PUBLICATION
3 December 2018PUBLISHED
8 January 2019Kyle J Lafata^{1,2,3,6} , Julian C Hong¹ , Ruiqi Geng^{1,2}, Bradley G Ackerson¹, Jian-Guo Liu^{3,4}, Zhennan Zhou⁵, Jordan Torok¹, Chris R Kelsey¹ and Fang-Fang Yin^{1,2,6}¹ Department of Radiation Oncology, Duke University Medical Center, Durham, NC 27710, United States of America² Medical Physics Graduate Program, Duke University, Durham, NC 27710, United States of America³ Department of Physics, Duke University, Durham, NC 27710, United States of America⁴ Department of Mathematics, Duke University, Durham, NC 27710, United States of America⁵ Beijing International Center of Mathematical Research, Peking University, Beijing, 100871, People's Republic of China⁶ Authors to whom any correspondence should be addressed.E-mail: kyle.lafata@duke.edu and fangfang.yin@duke.edu**Keywords:** radiomics, stereotactic body radiation therapy, non-small cell lung cancer, treatment outcomes**Abstract**

The purpose of this work was to investigate the potential relationship between radiomic features extracted from pre-treatment x-ray CT images and clinical outcomes following stereotactic body radiation therapy (SBRT) for non-small-cell lung cancer (NSCLC).

Seventy patients who received SBRT for stage-1 NSCLC were retrospectively identified. The tumor was contoured on pre-treatment free-breathing CT images, from which 43 quantitative radiomic features were extracted to collectively capture tumor morphology, intensity, fine-texture, and coarse-texture. Treatment failure was defined based on cancer recurrence, local cancer recurrence, and non-local cancer recurrence following SBRT. The univariate association between each radiomic feature and each clinical endpoint was analyzed using Welch's *t*-test, and *p*-values were corrected for multiple hypothesis testing. Multivariate associations were based on regularized logistic regression with a singular value decomposition to reduce the dimensionality of the radiomics data.

Two features demonstrated a statistically significant association with local failure: *Homogeneity2* ($p = 0.022$) and *Long-Run-High-Gray-Level-Emphasis* ($p = 0.048$). These results indicate that relatively dense tumors with a homogenous coarse texture might be linked to higher rates of local recurrence. Multivariable logistic regression models produced maximum AUC values of 0.72 ± 0.04 , 0.83 ± 0.03 , and 0.60 ± 0.04 , for the *recurrence*, *local recurrence*, and *non-local recurrence* endpoints, respectively.

The CT-based radiomic features used in this study may be more associated with local failure than non-local failure following SBRT for stage I NSCLC. This finding is supported by both univariate and multivariate analyses.

1. Introduction

Recent developments in computer science and technology have improved the feasibility of a data-driven interpretation of medical images. As such, quantitative imaging applications are quickly becoming standard-of-care in modern cancer medicine. Leading this paradigm shift is the emerging field of radiomics. Radiomics focuses on translating images into structured data to identify computational biomarkers (Kumar *et al* 2012, Aerts *et al* 2014, Gillies *et al* 2016). Computed tomography (CT) based radiomic analysis of non-small cell lung cancer (NSCLC) has been an attractive research topic in recent years (Aerts *et al* 2014, Huang *et al* 2016, Chaddad *et al* 2017, Chen *et al* 2017, Lee *et al* 2017, Timmeren *et al* 2017, Velden *et al* 2016). In particular, the association between radiomic features and stereotactic body radiation therapy (SBRT) outcomes has demonstrated promising results (Huynh *et al* 2016, Zhang *et al* 2016, Li *et al* 2017a, 2017b, Yu *et al* 2017, Oikonomou *et al* 2018). SBRT is quickly

becoming a standard treatment option for patients with early stage NSCLC, and therefore radiomics SBRT applications are highly relevant.

Radiomics has demonstrated prognostic value in predicting lung SBRT clinical outcomes, most notably for different aspects of patient survival (Huynh *et al* 2016, Zhang *et al* 2016, Li *et al* 2017a, 2017b, Yu *et al* 2017, Oikonomou *et al* 2018). However, fewer studies have focused on SBRT local failure rates. In part, this may be due to a generally low event-frequency of treatment failure in early stage cancers. Nevertheless, high local control has been a major rationale for SBRT. In fact, prior major prospective clinical trials have typically focused on this endpoint with respect to patterns of failure (Timmerman *et al* 2010, Chang *et al* 2012, Videtic *et al* 2015). Identifying radiomic features predictive of local failure may facilitate more aggressive therapy through: (a) modification of treatment volumes, (b) novel dose escalation techniques, and (c) closer follow-up for patients potentially at high risk of recurrence. Similarly, radiomic features predictive of non-local failures may help to identify patients at risk for metastatic disease progression who may benefit from adjuvant therapies after SBRT. In contrast to patient survival (which may be influenced by many different factors), radiomic prediction of local versus non-local failure may lead to more tangible implications within current SBRT workflows. The purpose of this study was to investigate the potential relationship between pre-treatment radiomic features and clinical outcomes, particularly local failure, following SBRT for NSCLC. Both univariate and multivariate analyses were performed.

2. Methods

2.1. Patient characteristics and clinical outcomes

This study was approved by the Institutional Review Board for radiomic analysis of stage I NSCLC patients treated with SBRT at Duke University between 2007 and 2014 ($n = 95$). Patients were excluded from the study if they: (a) had been previously treated for lung cancer ($n = 19$), or (b) presented with multiple synchronous primary cancers ($n = 6$). Baseline characteristics, treatment parameters, and follow-up clinical findings of the remaining 70 patients were retrospectively entered into a web-based application database. The median patient age was 74 years, there were 35 males to 35 females. Patients were treated to a mean dose of 51 Gy under a standard hypofractionation scheme. Patient-specific treatment outcomes were scored according to figure 1, based on follow-up CT, PET/CT, and pathological confirmation. The mean follow-up time was 65 months. Local treatment failure was confirmed pathologically in 33% of patients, and confirmed via PET/CT in 100% patients. Specifically, treatment outcomes were categorized as follows:

- **Failure** ($F \in \{0, 1\}$): Cancer recurrence following treatment ($n = 21$)
 - **Local failure** ($LF \in \{0, 1\}$): Cancer recurrence within 2 cm of the gross tumor volume (GTV) ($n = 6$)
 - **Non-local failure** ($nLF \in \{0, 1\}$): Either regional or distant failure ($n = 15$)
 - **Regional failure** ($RF \in \{0, 1\}$): Cancer recurrence within regional lymph nodes ($n = 5$)
 - **Distant failure** ($DF \in \{0, 1\}$): Development of metastatic disease ($n = 10$).

2.2. CT image acquisition and tumor contouring

For each patient, pre-treatment x-ray CT images were acquired under free-breathing conditions. Images were acquired on either a GE Lightspeed CT scanner (160 slices, 0.8 mm in plane resolution, 2.5 mm slice thickness, 120 kVp), or a Siemens Biograph mCT scanner (160 slices, 0.8 mm in plane resolution, 2 mm slice thickness, 120 kVp), and reconstructed using an FBP reconstruction algorithm. Both scanners were calibrated using an identical QA protocol and phantom. On each image, the GTV was manually segmented using commercially available contouring software. This was done on each axial CT slice, and the results were merged into a single 3D volume. All contours were verified by an experienced physician and physicist prior to radiomic feature extraction.

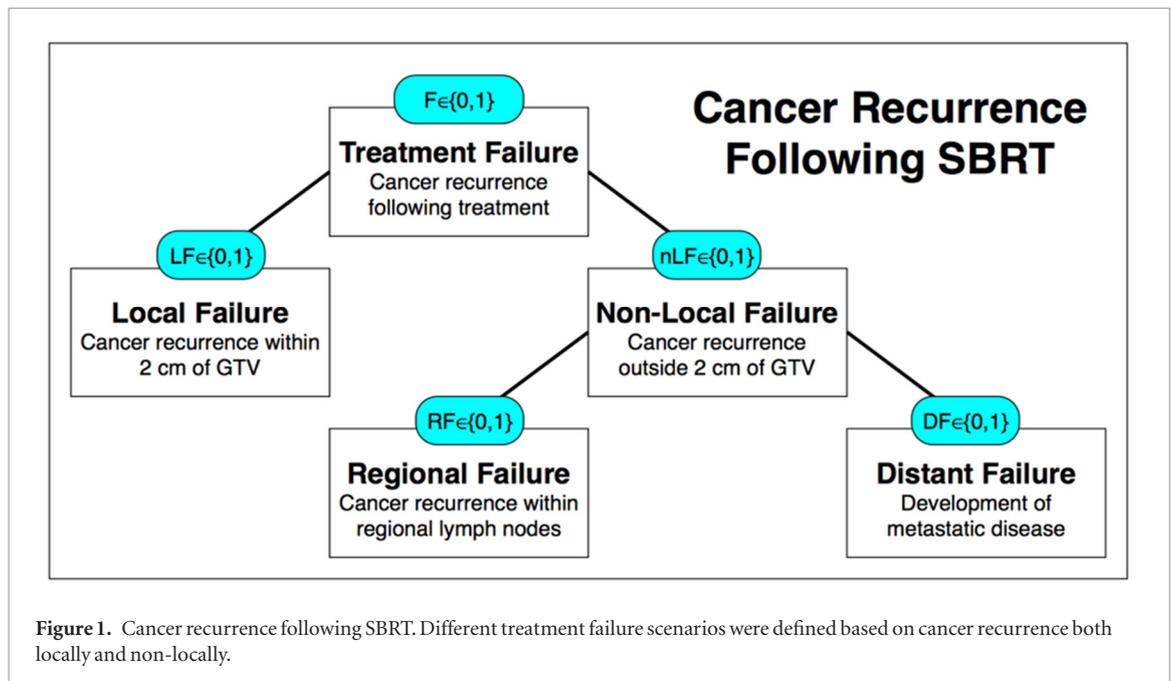
2.3. Radiomic feature extraction

Forty-three radiomic features were extracted from each GTV as potential biomarkers for NSCLC recurrence following SBRT. The resulting radiomic feature space was defined as the matrix

$$\mathcal{F} = (f_{ij}) \in \mathbb{R}^{43 \times 70}, \quad (1)$$

where, the (i, j) th coordinate of \mathcal{F} represents the measured value of the i th radiomic feature as observed within the image of the j th patient's GTV. For convenience, we define the j th patient's radiomic feature vector as,

$$\mathbf{f}_j = (f_{ij})_{i=1}^{43} \in \mathbb{R}^{43}, \quad j = 1, 2, \dots, 70. \quad (2)$$



Intensity		Fine Texture		Coarse Texture		Morphological					
#	Short	Feature Name	#	Short	Feature Name	#	Short	Feature Name			
1	I-1	Energy	3	F-1	Auto-correlation	27	C-1	Short Run Emphasis	38	M-1	Compactness 1
2	I-2	Entropy	6	F-2	Cluster Prominence	28	C-2	Long Run Emphasis	39	M-2	Compactness 2
3	I-3	Kurtosis	7	F-3	Cluster Shade	29	C-3	Gray Level Non-Uniformity	40	M-3	Sphericity
4	I-4	Skewness	8	F-4	Cluster Tendency	30	C-4	Run Length Non-Uniformity	41	M-4	Spherical Disproportion
			9	F-5	Contrast	31	C-5	Run Percentage	42	M-5	Surface Area
			10	F-6	Correlation	32	C-6	Low Gray Level Run Emphasis	43	M-6	Volume
			11	F-7	Differential Entropy	33	C-7	High Gray Level Run Emphasis			
			12	F-8	Dissimilarity	34	C-8	Short Run Low Gray Level Emphasis			
			13	F-9	GLCOM Energy	35	C-9	Short Run High Gray Level Emphasis			
			14	F-10	GLCOM Entropy	36	C-10	Long Run Low Gray Level Emphasis			
			15	F-11	Homogeneity1	37	C-11	Long Run High Gray Level Emphasis			
			16	F-12	Homogeneity2						
			17	F-13	Info Measure of Correlation 1						
			18	F-14	Info Measure of Correlation 2						
			19	F-15	Inverse Difference Moment Normalized						
			20	F-16	Inverse Difference Normalized						
			21	F-17	Inverse Variance						
			22	F-18	Maximum Probability						
			23	F-19	Sum Average						
			24	F-20	Sum Entropy						
			25	F-21	Sum Variance						
			26	F-22	Variance						

Figure 2. Radiomic features used in this analysis, color-coded by their class.

A complete list of all 43 features that were used to construct equation (2) is shown in figure 2. Briefly, each can be classified into one of the following 4 classes:

- **Intensity Features:** Measure the overall density characteristics of a tumor. These features were defined based on the image's gray level histogram. This is a probability density function (PDF) describing the frequency of gray-level occurrences in the image (Aerts *et al* 2014).
- **Fine Texture Features:** Capture small-scale heterogeneity within a tumor's detailed, high-resolution structure (i.e. at the imaging system's resolution-limit). These features were defined based on the image's Gray-Level Co-occurrence Matrix (Haralick *et al* 1973). By measuring the frequency of co-occurring adjacent voxel pairs having the same intensity, this joint PDF describes small scale spatial gray level dependencies.
- **Coarse Texture Features:** Capture large-scale heterogeneity within a tumor's approximate, low-resolution structure. These features were defined based on the image's Gray Level Run Length Matrix (Tang 1998). By measuring the frequency of *run-lengths* (i.e. the size of a set of consecutive voxels with the same grayscale intensity), this joint PDF describes large-scale spatial gray level dependencies.
- **Morphological Features:** Describe the overall 3D size and shape of a tumor.

As image texture is inherently scale-dependent, we chose to differentiate Fine Texture (small-scale) from Coarse Texture (large-scale). All texture calculations were performed in 13 directions corresponding to the axes-of-rotation of a cubic voxel. This was done to approximate a rotationally invariant 3D system. Feature extraction was performed in 3D using our in-house radiomics software developed in MATLAB (Mathworks, Natick, MA). Prior to analysis, this in-house feature extraction system was validated based on comparisons to the open source

software packages, CERR (Apte et al 2018) and IBEX (Zhang et al 2015). A complete list of features used in this study is presented in figure 2.

Given the 70 pairs of CT images and corresponding clinical outcomes, the following five datasets were constructed: A location unspecific failure dataset,

$$\left\{ \left(\mathbf{f}_j, y_j = F^{(j)} \right) \right\}_{j=1}^{70}, \quad (3)$$

as well as four location-specific sub-datasets,

$$\left\{ \left(\mathbf{f}_j, y_j = \text{LF}^{(j)} \right) \right\}_{j=1}^{70}, \quad (4)$$

$$\left\{ \left(\mathbf{f}_j, y_j = \text{RF}^{(j)} \right) \right\}_{j=1}^{70}, \quad (5)$$

$$\left\{ \left(\mathbf{f}_j, y_j = \text{DF}^{(j)} \right) \right\}_{j=1}^{70}, \quad (6)$$

$$\left\{ \left(\mathbf{f}_j, y_j = \text{nLF}^{(j)} \right) \right\}_{j=1}^{70}, \quad (7)$$

where, $\mathbf{f}_j \in \mathbb{R}^{43}$ is the j th patient's radiomic feature vector defined via equation (2), and $F^{(j)} \in \{0, 1\}$, $\text{LF}^{(j)} \in \{0, 1\}$, $\text{RF}^{(j)} \in \{0, 1\}$, $\text{DF}^{(j)} \in \{0, 1\}$, and $\text{nLF}^{(j)} \in \{0, 1\}$ are the previously defined clinical endpoints associated with the j th patient's cancer recurrence status.

2.4. Univariate statistical feature analysis

Based on equations (3)–(7), the statistical association between individual radiomic features and each clinical endpoint was analyzed using Welch's t -test (Welch 1947). All p -values were corrected for multiple hypothesis testing using the Bonferroni method (Bonferroni 1936). A corrected p -value ≤ 0.05 was considered statistically significant, indicating that the mean value of a particular feature's distribution could differentiate a given clinical endpoint.

2.5. Multivariate feature analysis

In this section, we investigate the multivariate association between the radiomics data and cancer recurrence using a regularized logistic regression model. As the number of features is relatively large compared to the number of patients, we chose to work in a truncated singular value decomposition (SVD) basis. SVD is a common matrix decomposition technique that is often used to reduce the dimensionality of a dataset by producing a set of uncorrelated, orthonormal bases. In the analyses that follow, each clinical endpoint (i.e. $F \in \{0, 1\}$, $\text{LF} \in \{0, 1\}$, $\text{RF} \in \{0, 1\}$, $\text{DF} \in \{0, 1\}$, and $\text{nLF} \in \{0, 1\}$) was treated independently as a separate analysis.

2.5.1. Singular value decomposition

Each column vector of \mathcal{F} was zero-mean centered to ensure normalized units. An SVD transformation was subsequently performed on the feature space,

$$\mathcal{F} = \sum_{i=1}^k \sigma_i \mathbf{u}_i \mathbf{v}_i^T, \quad (8)$$

where, \mathbf{u}_i and \mathbf{v}_i are the left-and-right singular vectors of \mathcal{F} , respectively, and σ_i are the singular values of \mathcal{F} . The vectors, \mathbf{u}_i , which are similarly known as Principal Components, provide a unique k -dimensional coordinate system for the data. Equation (8) was truncated to the first k leading terms to define a lower-dimensional feature space in the Principal coordinate axes, $(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k)$. The j th patient's radiomic feature vector was defined in this basis as,

$$\mathbf{u}_j = \left(u_{ij} \right)_{i=1}^n \in \mathbb{R}^n, \quad j = 1, 2, \dots, k; \quad k < n = 43. \quad (9)$$

2.5.2. Multivariate logistic regression

For each \mathbf{u}_j , the following logistic function was defined

$$g_j = \varphi \left(w_o + \mathbf{w}^T \mathbf{u}_j \right), \quad j = 1, 2, \dots, k, \quad (10)$$

where,

$$\varphi(x) = \frac{1}{1 + e^{-x}}, \quad (11)$$

and \mathbf{w} is a weight vector with bias term, $w_0 = 1$. Given m input/output training pairs, \mathbf{w} was optimized by minimizing a regularized cross-entropy loss function (Murphy 2012),

$$J(\mathbf{w}) = \operatorname{argmin} -\frac{1}{k} \sum_{j=1}^k y_j \log [g_j] + (1 - y_j) \log [1 - g_j] + \lambda \sum_{i=1}^n w_i^2 \quad (12)$$

where, g_j is the modeled response to the j th input according to equation (10), y_j is the ground-truth class associated with the j th patient according to equations (3)–(7), and $\lambda > 0$ is a regularization parameter. The parameter λ defines the number of non-zero components of \mathbf{w} , and was determined during the training phase via cross-validation. Equation (12) was minimized as a batch process across the set of m input/output training pairs, using a standard conjugate gradient algorithm.

2.5.3. Model performance evaluation

Model generalization was evaluated using a stratified Monte Carlo cross-validation approach with 25 iterations (Burman 1989). At each iteration, equations (3)–(7) were each randomly partitioned into training (80%) and validation (20%) cohorts with approximately equal event ratios. Models were developed using only the training cohort based on equations (8)–(12), and were applied to the validation cohort. Results were pooled to construct receiver operating characteristic (ROC) curves, and the area under the curve (AUC) was used as the final model performance metric. To compare different AUC values, pair-wise p -values were calculated based on a permutation test with 250 iterations. Finally, as a quality check, the above methodology was performed on data classes that were artificially re-assigned a random binary value. In theory, such a scenario should result in $\text{AUC} = 0.5$. This step was implemented as a benchmark control to ensure that models were not grossly over-fit to the training data. In practice, modeling such a random distribution should result in $\text{AUC} \approx 0.5$ (due to the noise in the data).

3. Results

3.1. Univariate feature analysis

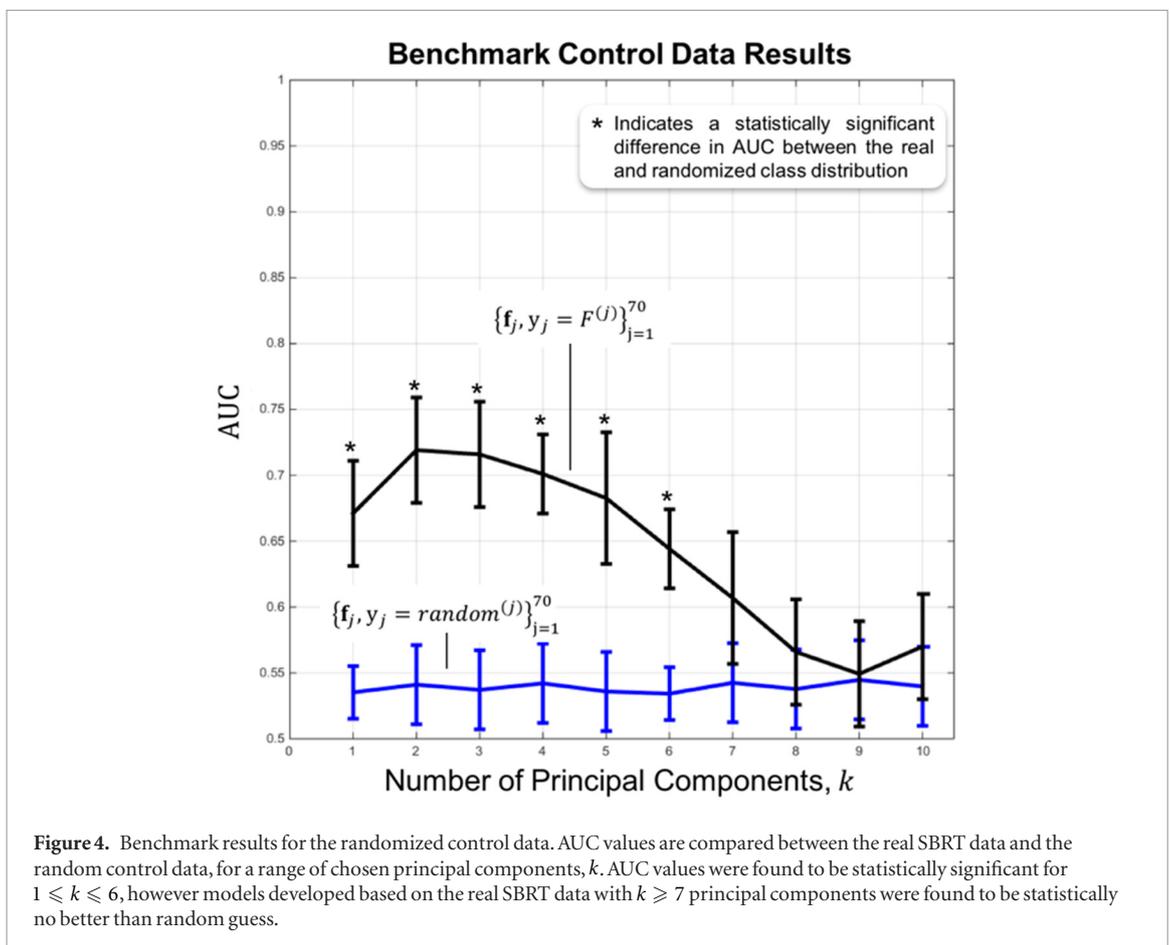
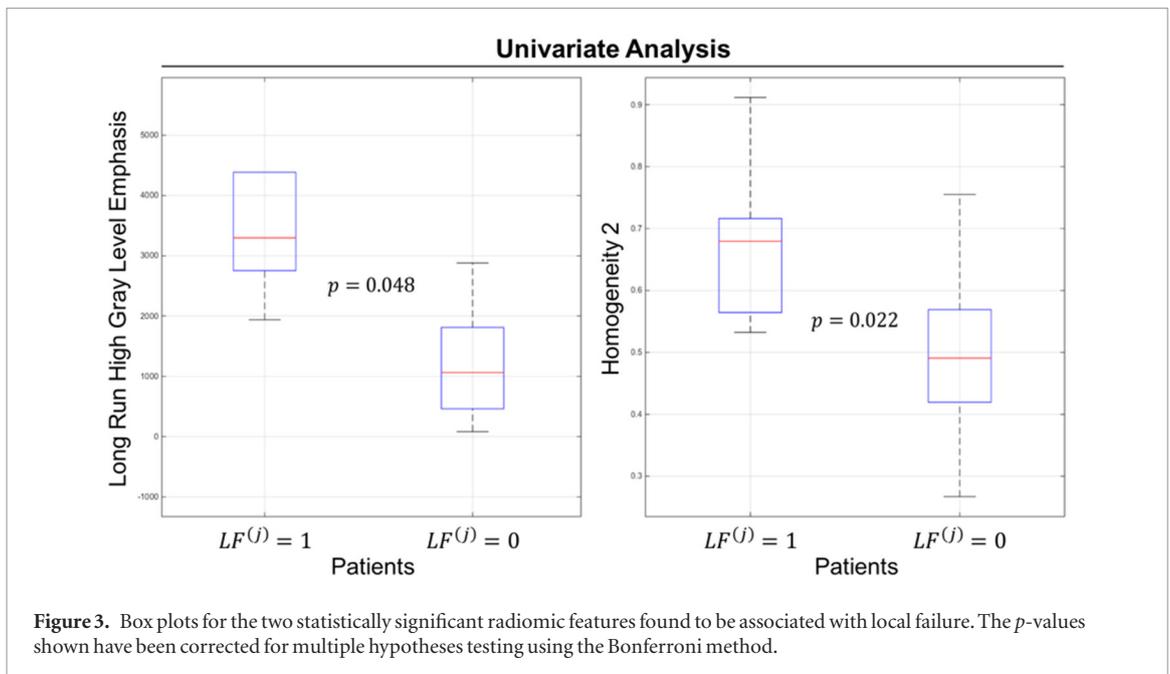
Two radiomic features demonstrated a statistically significant univariate association with $\text{LF}^{(+)}$: *Homogeneity 2* ($p = 0.022$) and *Long-Run-High-Gray-Level-Emphasis* ($p = 0.048$). The former is a fine texture feature, while the latter is a coarse texture feature. These results indicate that relatively dense tumors with a homogenous coarse texture were often linked to higher rates of local recurrence. Box plots for each analysis is shown in figure 3. No such relationship was observed for the $\text{F}^{(+)}$, $\text{RF}^{(+)}$, $\text{DF}^{(+)}$, or $\text{nLF}^{(+)}$ cohorts, where all p -values lost their statistical significance following the Bonferroni correction.

3.2. Multivariate feature analysis

Figure 4 summarizes the benchmark control results based on data classes that were artificially re-assigned a random binary value. The AUC of the randomized control data was close to the theoretical 0.5, and was found to be stable for all Principal Component cutoffs, k . When comparing the *real* SBRT data to the *random* control data, AUC values were found to be statistically significant for $1 \leq k \leq 6$. Models developed based on the real SBRT data with ($k \geq 7$) principal components were found to be statistically no better than random guess.

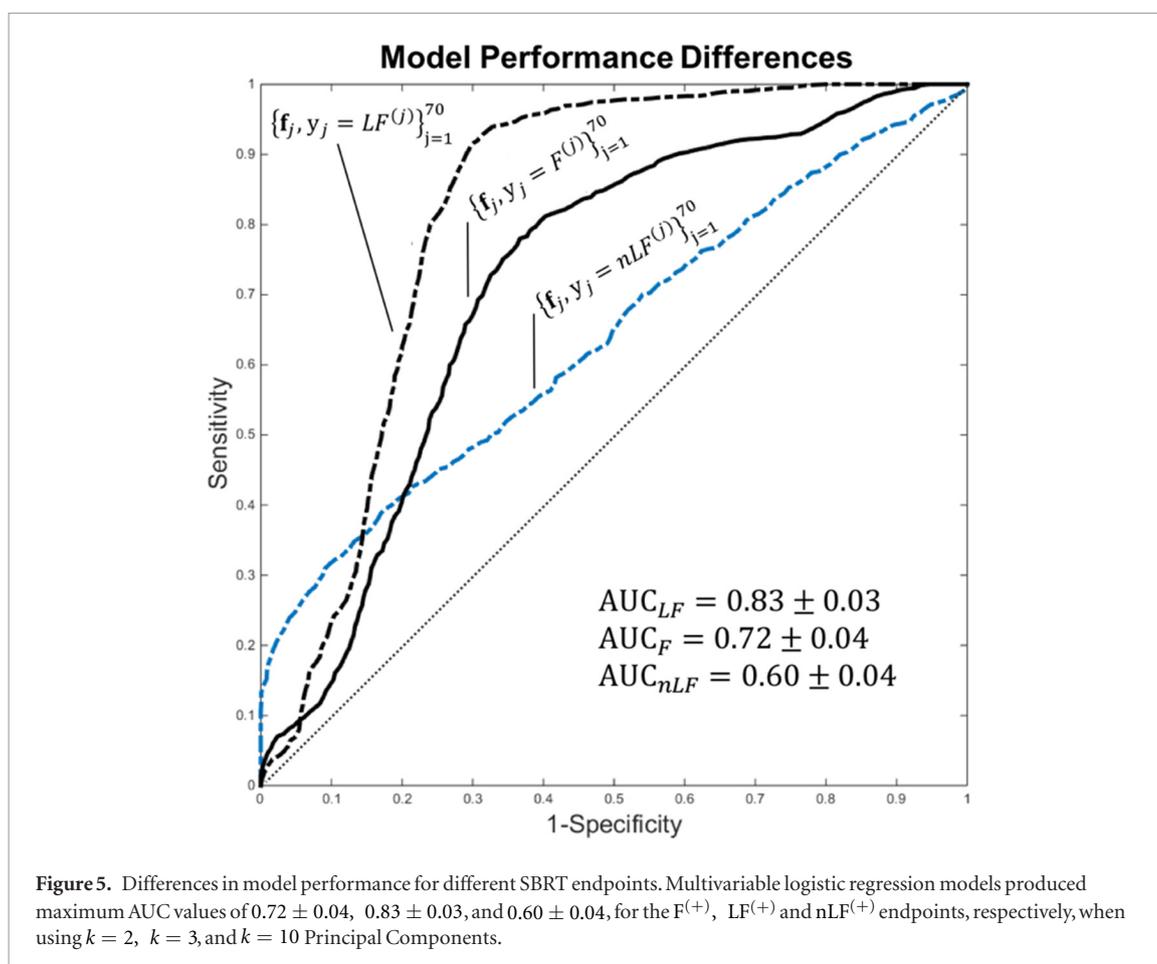
Multivariable logistic regression models produced maximum AUC values of 0.72 ± 0.04 , 0.83 ± 0.03 , and 0.60 ± 0.04 , for the $\text{F}^{(+)}$, $\text{LF}^{(+)}$, and $\text{nLF}^{(+)}$ endpoints, respectively. These results are demonstrated in figure 5, and were produced with $k = 2$, $k = 3$, and $k = 10$ Principal Components, respectively. None of the $\text{RF}^{(+)}$ and $\text{DF}^{(+)}$ models were found to be statistically significant when compared to the combined $\text{nLF}^{(+)}$ model.

As summarized in figure 6, performance was sensitive to the number of Principle Components used in each model. A statistically significant AUC difference was found between the $\text{F}^{(+)}$, $\text{LF}^{(+)}$, and $\text{nLF}^{(+)}$ models for $(1 < k < 6)$ Principal Components. However, when $k \geq 6$, AUC was found to be statistically insignificant between all models. In general, the performance of both the failure and local failure models is shown to decrease as $(k > 5) \rightarrow 10$.



4. Discussion

Preliminary results shown in figures 3–6 indicate that radiomics has the potential to be used for personalized medicine. Applications of radiomics to SBRT are particularly intriguing, since SBRT is being increasingly considered as a standard treatment option for medically inoperable NSCLC patients with early stage disease. The ability to identify radiomic biomarkers from pre-SBRT CT images may provide additional key information that is readily accessible during treatment planning. This may potentially allow for more aggressive therapy or modification of treatment strategies. In general, our study suggests that CT radiomics may carry more predictive



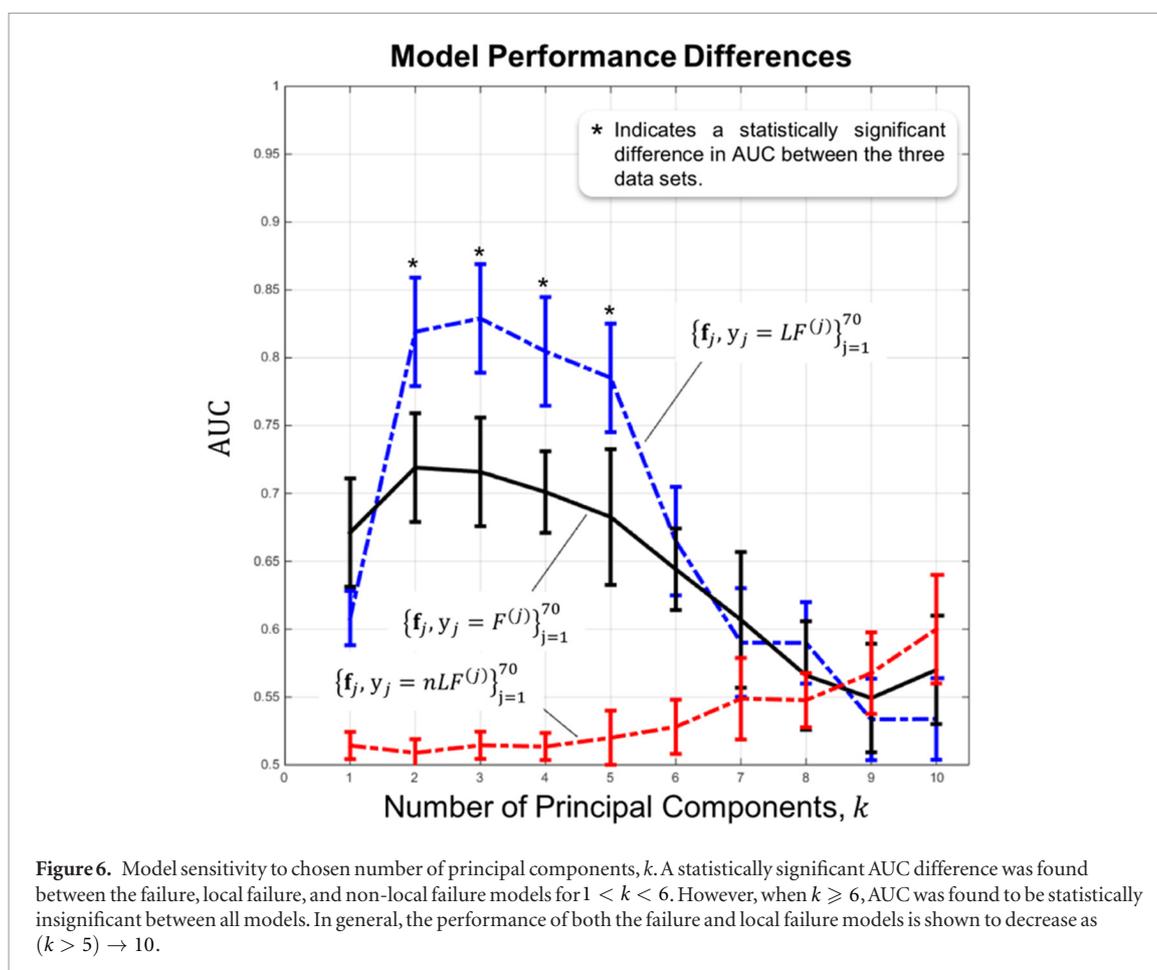
power in the identification of local SBRT failures, rather than non-local failures. This finding is supported by both univariate and multivariate analyses.

Huynh *et al* (2016) first investigated the prognostic value of pre-SBRT CT radiomic features. They evaluated overall survival (OS), distant metastasis (DM), and locoregional recurrence (LRR) as clinical endpoints for stage I–II NSCLC patients, and concluded that radiomic features may be significantly more predictive than conventional imaging metrics (e.g. tumor volume). Li *et al* (2017a) similarly found radiomics data derived from planning-CT images demonstrated prognostic value for recurrence following SBRT. Their study included stage I–II NSCLC patients and investigated OS, recurrence-free survival (RFS), and locoregional recurrence-free survival (LR-RFS). Their findings indicated that statistical modeling of these endpoints based on clinical data generally improved with the addition of radiomic features. Our results are consistent with both of these studies, as we found cancer recurrence to be more associated with radiomics data, rather than conventional imaging metrics, such as tumor volume.

Following our univariate analysis, *Homogeneity 2* and *Long-Run-High-Gray-Level-Emphasis* were both found to be statistically significant in differentiating the $LF^{(+)}$ cohort from the $LF^{(-)}$ cohort (figure 3). None of the radiomic features were able to differentiate $F^{(+)}$ from $F^{(-)}$, $RF^{(+)}$ from $RF^{(-)}$, $DF^{(+)}$ from $DF^{(-)}$, or $nLF^{(+)}$ from $nLF^{(-)}$. These results are reported in figure 3, and imply that the $LF^{(+)}$ tumors were on average more dense and homogenous than their $LF^{(-)}$ counterparts.

Overall, there were no obvious pre-treatment patterns within the $LF^{(+)}$ cohort. In fact, these patients seemed to be fairly diverse in many non-radiomic factors, including, tumor histology, tumor size, tumor location, and fractionation scheme. This trend is consistent with other published studies that have reported worse predictive power when modeling SBRT failures based solely on conventional imaging metrics (Huynh *et al* 2016) or clinical data (Li *et al* 2017a).

Interestingly, Yu *et al* (2017) had also found that *Homogeneity 2* was associated with SBRT mortality rates for early stage NSCLC. They developed and validated a CT-based NSCLC radiomic signature by training on a surgical cohort and testing on an SBRT cohort. In this study, *Homogeneity 2* was found to be an independent predictor of overall survival (Yu *et al* 2017). Further, a denser tumor may potentially possess a higher cell count, which would imply a larger, more difficult-to-treat tumor burden. A study by Ye *et al* (2015) supports this claim. They found denser tumors to be more likely associated with poorer SBRT outcomes, and reported tumors greater than 0.7 g cm^{-3} to be correlated with inferior disease free survival rates (Ye *et al* 2015).



Our multivariate results were consistent with the univariate analysis discussed above. As demonstrated in figure 4, model performance for $F^{(+)}$, $LF^{(+)}$, and $nLF^{(+)}$ was maximized at $AUC = 0.72 \pm 0.04$, $AUC = 0.83 \pm 0.03$, and $AUC = 0.60 \pm 0.04$, respectively. These metrics were produced at $k = 2$, $k = 3$, and $k = 10$ Principal Components, respectively, and were found to be statistically significant from each other. It is noted that none of the $RF^{(+)}$ and $DF^{(+)}$ models were found to be statistically significant when compared to the combined $nLF^{(+)}$ models. This indicates that modeling location-specific, non-local recurrence was not feasible with our dataset.

As shown in figure 5, the $LF^{(+)}$ model performed better than all other models for ($2 \leq k \leq 5$) Principal Components. In this range, the mean $LF^{(+)}$ AUC was found to be 0.81, compared to 0.72 and 0.53 for the $F^{(+)}$ and $nLF^{(+)}$ models, respectively. In general, however, the performance of both the $LF^{(+)}$ and $F^{(+)}$ models was found to decrease as ($k > 5$) \rightarrow 43. When ($6 \leq k \leq 43$) was used, $LF^{(+)}$ and $F^{(+)}$ AUC values were found to be statistically indistinguishable from $nLF^{(+)}$ counterparts. This is likely a result of having too many degrees-of-freedom, and intuitively makes sense.

In contrast to the $F^{(+)}$ and $LF^{(+)}$ results, we were unable to generate strong logistic regression models for the $RF^{(+)}$, $DF^{(+)}$, and $nLF^{(+)}$ endpoints. The best-performing $nLF^{(+)}$ model achieved $AUC = 0.60 \pm 0.04$, which was significantly worse than the corresponding $F^{(+)}$ and $LF^{(+)}$ models. Overall, these multivariate findings—in conjunction with univariate results—suggest that our chosen radiomic features may not be as strongly correlated with the underlying mechanisms that are responsible for disease spread. However, these conclusions remain limited to be tested with a larger sample size. In particular, we note that while this paper demonstrates strong feasibility of using radiomics to predict local cancer recurrences following SBRT, a limitation of the work is the lack of an external test dataset. Further evaluation of the technique using a separate test dataset is the focus of our future work. This will provide a more robust assessment of model generalization.

Finally, our multivariate methodology consisted of combining two relatively common feature processing techniques: singular value decomposition (SVD) and LASSO regularization. The rationale here was as follows. Without loss of generality, we chose to work in a truncated SVD basis to reduce the overall degrees-of-freedom of the data. This is important as the original data is of high dimension with a relatively small number of samples. However, SVD is simply a change-of-basis and therefore likely will not be adequate for feature-selection. Further, it is reasonable to hypothesize that some Principal Components are more important than others when considering a specific target variable. This is where the LASSO operator is beneficial. By implementing regularization in conjunction with SVD, the LASSO operator has fewer weight components to optimize during the training pro-

cess. We note that D'Angelo *et al* (2009) successfully used a similar approach to identify gene-gene interactions. Our findings indicate that the hybrid SVD-LASSO approach improved model performance and generalization.

5. Conclusions

For stage I NSCLC, the CT-based radiomic features used in this study may be more predictive of local SBRT failure than non-local SBRT failure. This finding is supported by both univariate and multivariate analyses. *Homogeneity 2* and *Long-Run-High-Gray-Level-Emphasis* were found to be statistically associated with local cancer recurrence. Further, multivariable logistic regression models demonstrated better performance and generalization for local failure prediction, relative to non-local failure prediction.

ORCID iDs

Kyle J Lafata  <https://orcid.org/0000-0002-4513-6249>

Julian C Hong  <https://orcid.org/0000-0001-5172-6889>

References

- Aerts H *et al* 2014 Decoding tumor phenotype by noninvasive imaging using a quantitative radiomics approach *Nat. Commun.* **5** 1–8
- Apte A *et al* 2018 Technical note: extension of CERR for computational radiomics: a comprehensive MATLAB platform for reproducible radiomics research *Med. Phys.* (<https://doi.org/10.1002/mp.13046>)
- Bonferroni C 1936 Teoria statistica delle classi e calcolo delle probabilità, Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze
- Burman P 1989 A Comparative study of ordinary cross-validation, v -fold cross validation and the repeated learning testing-model methods *Bometrika* **76** 503–14
- Chaddad A *et al* 2017 Predicting survival time of lung cancer patients using radiomic analysis *Oncotarget* **8** 104393–407
- Chang J *et al* 2012 Clinical outcome and predictors of survival and pneumonitis after stereotactic ablative radiotherapy for stage I non-small cell lung cancer *Radiat. Oncol.* **7** 152
- Chen B *et al* 2017 Development and clinical application of radiomics in lung cancer *Radiat. Oncol.* **12** 154
- D'Angelo X *et al* 2009 Combining least absolute shrinkage and selection operator (LASSO) and principal-components analysis for detection of gene-gene interactions in genome-wide association studies *BMC Proc.* **3**(Suppl. 7) S62
- Gillies R *et al* 2016 Radiomics: images are more than pictures, they are data *Radiology* **278** 563–77
- Haralick R 1973 Texture features for image classification *IEEE Trans. Syst. Man Cybern.* **6** 610–21
- Huang Y *et al* 2016 Radiomics signature: a potential biomarker for the prediction of disease-free survival in early-stage (I or II) non-small cell lung cancer *Radiology* **281** 947–57
- Huynh E *et al* 2016 CT-based radiomic analysis of stereotactic body radiation therapy patients with lung cancer *Radiother. Oncol.* **120** 258–66
- Kumar V *et al* 2012 Radiomics: the process and the challenges *Magn. Reson. Imaging* **30** 1234–48
- Lee G *et al* 2017 Radiomics and its emerging role in lung cancer research, imaging biomarkers and clinical management: state of the art *Eur. J. Radiol.* **86** 297–307
- Li Q *et al* 2017a Imaging features from pretreatment CT scans associated with clinical outcomes in nonsmall-cell lung cancer patients treated with stereotactic body radiotherapy *Med. Phys.* **44** 4341–9
- Li Q *et al* 2017b CT imaging features associated with recurrence in non-small cell lung cancer patients after stereotactic body radiotherapy *Radiother. Oncol.* **12**
- Murphy K 2012 *Machine Learning: a Probabilistic Perspective* (Cambridge: The MIT Press)
- Oikonomou A *et al* 2018 Radiomics analysis at PET/CT contributes to prognosis of recurrence and survival in lung cancer treated with stereotactic body radiotherapy *Sci. Rep.* **8** 4003
- Tang X 1998 Texture information in run length matrices *IEEE Trans. Image Process.* **7** 1602–9
- Timmeren R *et al* 2017 Survival prediction of non-small cell lung cancer patients using radiomics analyses of cone-beam CT images *Radiother. Oncol.* **124** 363–9
- Timmerman R *et al* 2010 Stereotactic body radiation therapy for inoperable early stage lung cancer *JAMA* **303** 1070–6
- Velden F *et al* 2016 Repeatability of radiomic features in non-small-cell lung cancer F-18 FDG-PET/CT studies: impact of reconstruction and delineation *Mol. Imaging Biol.* **18** 788–95
- Videtic G *et al* 2015 A randomized phase 2 study comparing two stereotactic body radiation therapy schedules for medically inoperable patients with stage I peripheral non-small cell lung cancer: NRG Oncology RTOG 0915 (NCCTG N0927) *Int. J. Radiat. Oncol. Biol. Phys.* **93** 757–64
- Welch B 1947 The generalization of 'student's' problem when several different population variances are involved *Biometrika* **34** 28–35
- Ye J *et al* 2015 Tumor density, size, and histology in the outcome of stereotactic body radiation therapy for early-stage non-small-cell lung cancer: a single-institution experience *Proc. of the 97th Annual Meeting of the American Radium Society*
- Yu W *et al* 2017 Development and validation of a predictive radiomics model for clinical outcomes in stage I non-small cell lung cancer *Int. J. Radiat. Oncol. Biol. Phys.* **102** 1090–7
- Zhang L *et al* 2015 IBEX: an open infrastructure software platform to facilitate collaborative work in radiomics *Med. Phys.* **42** 1341–53
- Zhang Y *et al* 2016 Radiomics-based prognosis analysis for non-small cell lung cancer *Sci. Rep.* **7** 46349