

# Radial gradient and radial deviation radiomic features from pre-surgical CT scans are associated with survival among lung adenocarcinoma patients

Ilke Tunali<sup>1,3,4</sup>, Olya Stringfield<sup>1</sup>, Albert Guvenis<sup>3</sup>, Hua Wang<sup>5</sup>, Ying Liu<sup>5</sup>, Yoganand Balagurunathan<sup>1</sup>, Philippe Lambin<sup>6</sup>, Robert J. Gillies<sup>1</sup> and Matthew B. Schabath<sup>2</sup>

<sup>1</sup>Department of Cancer Imaging and Metabolism, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida, USA

<sup>2</sup>Department of Cancer Epidemiology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida, USA

<sup>3</sup>Institute of Biomedical Engineering, Bogazici University, Istanbul, Turkey

<sup>4</sup>Faculty of Biomedical Engineering, Namik Kemal University, Tekirdag, Turkey

<sup>5</sup>Department of Radiology, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center of Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin, PR China

<sup>6</sup>Research Institute GROW of Oncology, Maastricht University Medical Center, Maastricht, The Netherlands

**Correspondence to:** Matthew B. Schabath, **email:** Matthew.Schabath@moffitt.org

**Keywords:** radiomics; radial gradient; radial deviation; lung adenocarcinoma; quantitative imaging

**Received:** March 12, 2017

**Accepted:** August 26, 2017

**Published:** October 06, 2017

**Copyright:** Tunali et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

**The goal of this study was to extract features from radial deviation and radial gradient maps which were derived from thoracic CT scans of patients diagnosed with lung adenocarcinoma and assess whether these features are associated with overall survival. We used two independent cohorts from different institutions for training (n= 61) and test (n= 47) and focused our analyses on features that were non-redundant and highly reproducible. To reduce the number of features and covariates into a single parsimonious model, a backward elimination approach was applied. Out of 48 features that were extracted, 31 were eliminated because they were not reproducible or were redundant. We considered 17 features for statistical analysis and identified a final model containing the two most highly informative features that were associated with lung cancer survival. One of the two features, radial deviation outside-border separation standard deviation, was replicated in a test cohort exhibiting a statistically significant association with lung cancer survival (multivariable hazard ratio = 0.40; 95% confidence interval 0.17-0.97). Additionally, we explored the biological underpinnings of these features and found radial gradient and radial deviation image features were significantly associated with semantic radiological features.**

## INTRODUCTION

Lung cancer is the second most common cancer and is the leading cause of cancer-related death in the United States. Lung cancer accounts for more deaths than prostate, breast, colon, and pancreatic cancer combined [1]. Despite improvements in survival for many other cancer types over the last several decades, there has been little improvement in lung cancer patient survival, mainly because of the fact that by the time a diagnosis is made, the

cancer is often in advanced stages and treatment options are limited. The five-year survival rate for all lung cancers (non-small cell lung carcinoma [NSCLC] and small cell lung cancer combined) is only 17%; and among NSCLC diagnoses, the five-year relative survival rate is 21% [2].

Pathologic staging is the most important prognostic factor for lung cancer survival [3]. However, there is marked variability in patient outcomes and survival among patients with the same stage of disease, which suggests that other factors contribute to NSCLC prognosis.

These prognostic factors include sex, histology, genetic alterations in oncogenes and tumor suppressor genes, co-morbidities, and patient performance status [4–9]. Additionally, there is emerging evidence that radiological and quantitative image features are associated with patient outcomes independent of clinical covariates and patient characteristics [2, 10–14]. As diagnostic computed tomography (CT) scans are routinely obtained during the workup of lung cancer patients, image features can provide valuable and readily available complementary decision support information which could have translational implications for improved prediction of patient outcomes and further patient stratification.

With high-throughput computing, it is now possible to rapidly extract a large number of quantitative image features from standard-of-care imaging such as CT. The conversion of digital medical images into mineable high-dimensional data is a process that is known as radiomics. Radiomics is motivated by the premise that biomedical images contain information that reflects the underlying pathophysiology of the region of interest (i.e., lung tumor) and that these relationships can be revealed via conversion of images to structured data, data-mining, and statistical analysis [15]. In this study, we analyzed a set of image features extracted from radial gradient (RG) and radial deviation (RD) maps generated from thoracic CT images. For each voxel in the volume of interest (VOI), a radial deviation and a radial gradient value was calculated which in-turn formed the radial gradient and radial deviation maps. Each voxel in the radial deviation map is defined as the angle between a voxel's gradient vector and its radial vector which points towards the center of mass of the segmented lesion, whereas each voxel in the radial gradient map specified the magnitude of gradient along that voxel's radial vector. Using these maps, we generated radial gradient and radial deviation features which represent voxel-by-voxel gradient changes in the VOI. As such, we expected that these features will be sensitive to changes in tumor shape that occur along radial directions, such as lobulation and border definition, which are important predictive and prognostic features in lung cancer [11, 16, 17].

To date there have been very few studies on image features derived from radial gradient and radial deviation maps [18–20]. Features extracted from radial gradient and radial deviation maps were first used in a computer-aided detection (CAD<sub>d</sub>) system for eliminating false positive pulmonary nodule candidates on chest X-ray [18]. Messay et al. [19] used these features in a computer-aided diagnosis system (CAD<sub>x</sub>) to discriminate between benign and malignant nodules. In another study from this group, radial gradient and radial deviation image features were utilized to optimize free parameters of a CT pulmonary nodule segmentation system [20]. As such, the goal of this study was to extract features from radial deviation and radial gradient maps from pre-surgical

contrast-enhanced thoracic CT scans among patients with lung adenocarcinoma and assess whether these features were associated with clinical outcomes. Additionally, we explored the potential biological underpinnings of these features by analyzing the association between radial gradient and radial deviation image features with semantic radiological features.

## RESULTS

### Patient demographics

Among the 61 patients in the training cohort, 50.8% were male, 67.2% were aged above 65 years at the date of diagnosis, and 72.1% were either stage I or II. Among the image acquisition parameters, 93.4% of the CT scans were acquired with 120 kVp, 34.4% used B41f as a convolution kernel, 65.6% had an interpolated slice thickness of 2.5 mm, and 34.4% had a pixel resolution  $\geq 0.7785$  (third quartile). The median time to event (overall survival) was 33.5 months for this cohort (Table 1).

In test cohort, there were a total of 47 patients of which 53.2% were male, 46.8% were aged above 65 years at the date of diagnosis, and 68.1% were either stage I or II. Among the image acquisition parameters, 85.1% of the CT scans were acquired with 120 kVp, 48.9% used A or B as a convolution kernel, 61.7% had an interpolated slice thickness of 2.5 mm, and 78.7% had a pixel resolution  $\geq 0.7785$  (third quartile). The median time to event (overall survival) was 32.0 months for this cohort (Table 1).

### Univariable analyses

After eliminating the redundant ( $n = 15$ ) and non-reproducible features ( $n = 16$ ), we calculated the log-rank p-values for the remaining 17 features. Out of these 17 features, two features (radial deviation outside-border separation standard deviation (SD) and radial gradient outside-border separation SD 2-dimensional (2D)) were statistically significantly associated with overall survival, (log-rank p-value  $\leq 0.05$ ) and three features (radial gradient border SD, radial gradient outside-tumor separation mean, and radial deviation tumor SD) were marginally (log-rank p-value  $\leq 0.1$ ) associated with overall survival (Table 2). The log-rank p-values for all 17 features assessed are presented in Supplementary Table 1.

The Kaplan-Meier survival curves using a median cutoff for the five features are presented in Figure 1A to 1e and the 5-year survival rates are presented in Supplementary Table 2. For the two features that were significantly associated with overall survival, tumors with high ( $\geq$  median) radial deviation outside-border separation SD (Figure 1C, Hazard Ratio [HR] = 0.36; 95% CI 0.16-0.81,  $p = 0.013$ ) and radial gradient outside-border separation SD (Figure 1D, HR = 0.43; 95% CI 0.20-0.94,

**Table 1: Patient characteristics in the training and test cohorts**

Characteristic	Training cohort (N = 61)		Test cohort (N = 47)	
<b>Age at diagnosis, N (%)</b>				
< 65	20	(32.8)	25	(53.2)
≥ 65	41	(67.2)	22	(46.8)
<b>Sex, N (%)</b>				
Female	30	(49.2)	22	(46.8)
Male	31	(50.8)	25	(53.2)
<b>Stage, N (%)</b>				
I and II	44	(72.1)	32	(68.1)
III and IV	17	(27.9)	15	(31.9)
<b>Tumor volume, mean cm<sup>3</sup> (SD)</b>	19.5	(29.0)	52.4	(130.0)
<b>Tumor max diameter, mean mm (SD)</b>	31.6	(13.8)	38.0	(21.5)
<b>Overall Survival, median months</b>	33.5		32.0	
<b>Image acquisition parameters, N (%)</b>				
<i>Voltage, kVp</i>				
120	57	(93.4)	40	(85.1)
130 or 140	4	(6.6)	7	(14.9)
<i>Convolution kernel</i>				
A,B	0	(0)	23	(48.9)
B30s,B60f,B70s	2	(3.3)	5	(10.7)
B30f	8	(13.1)	0	(0)
B40f	19	(31.2)	15	(31.9)
B41f	21	(34.4)	0	(0)
Other	11	(18.0)	4	(8.5)
<i>Interpolated slice thickness</i>				
1.5 mm	0	(0)	2	(4.3)
2.0 mm	8	(13.1)	13	(27.7)
2.5 mm	40	(65.6)	29	(61.7)
3.0 mm	13	(21.3)	3	(6.3)
<i>Pixel resolution (mm), tertiles</i>				
< 0.6926	20	(32.8)	6	(12.8)
≥ 0.6926 to < 0.7785	20	(32.8)	4	(8.5)
≥ 0.7785	21	(34.4)	37	(78.7)

**Table 2: Log-rank tests and Cox proportional hazards model for overall survival in the training and test cohorts**

Covariate	Training cohort N = 61						Test cohort N = 47				
	Log-rank P-value <sup>1</sup>	Univariable model <sup>2</sup> OR (95% CI)	P-value	Multivariable model <sup>3</sup> OR (95% CI)	P-value	Multivariable model <sup>4</sup> OR (95% CI)	P-value	Multivariable model <sup>5</sup> OR (95% CI)	P-value	Multivariable model <sup>6</sup> OR (95% CI)	P-value
Radial gradient border SD (feature 20)	0.084	1.92 (0.90 - 4.11)	0.092	.	.	.	.	.	.	.	.
Radial gradient outside-tumor separation mean (feature 35)	0.061	0.48 (0.22 - 1.06)	0.068	<b>0.29 (0.12 - 0.66)</b>	<b>0.003</b>	<b>0.31 (0.13 - 0.72)</b>	<b>0.006</b>	0.75 (0.28 - 2.03)	0.575	0.48 (0.17 - 1.37)	0.172
Radial deviation outside-border separation SD (feature 42)	<b>0.009</b>	<b>0.36 (0.16 - 0.81)</b>	<b>0.013</b>	<b>0.25 (0.11 - 0.58)</b>	<b>0.001</b>	<b>0.24 (0.10 - 0.58)</b>	<b>0.001</b>	<b>0.36 (0.16 - 0.81)</b>	<b>0.014</b>	<b>0.40 (0.17 - 0.97)</b>	<b>0.042</b>
Radial gradient outside-border separation SD (2D) (feature 48)	<b>0.029</b>	<b>0.43 (0.20 - 0.94)</b>	<b>0.035</b>	.	.	.	.	.	.	.	.
Radial deviation tumor SD (feature 2)	0.071	2.00 (0.92 - 4.34)	0.078	.	.	.	.	.	.	.	.
Age	0.439	1.38 (0.60 - 3.16)	0.444	.	.	0.83 (0.34 - 2.05)	0.690	.	.	<b>2.65 (1.07 - 6.60)</b>	<b>0.035</b>
Sex	0.694	1.16 (0.54 - 2.49)	0.696	.	.	1.05 (0.47 - 2.35)	0.906	.	.	1.43 (0.53 - 3.82)	0.476
Stage	0.085	1.95 (0.90 - 4.23)	0.093	.	.	2.14 (0.91 - 5.03)	0.082	.	.	<b>3.35 (1.34 - 8.36)</b>	<b>0.010</b>
Tumor volume	<b>0.044</b>	2.23 (1.00 - 4.97)	0.051	.	.	.	.	.	.	.	.

SD = standard deviation; OR = odds ratio; CI = confidence interval  
 Bold values are statistically significant.

<sup>1</sup>Log-rank p-value for each covariate for overall survival right censored at 5-years. The radiomic features were dichotomized at the median value and the clinical covariates were dichotomized based on Table 1. The univariable analyses were based on 62 patients. But, due to missing patient data (age and sex), the total sample size for the multivariable analyses was 61 patients.

<sup>2</sup>The independent main effect ORs for each covariate

<sup>3</sup>The ORs for the two image features in a single model following backward elimination that considered all features and tumor volume.

<sup>4</sup>The ORs for both image features identified from backward elimination adjusted for clinical covariates.

<sup>5</sup>The ORs for from the two image features identified in training cohort from backward elimination

<sup>6</sup>The ORs for both image features identified from backward elimination in training cohort adjusted for clinical covariates

p = 0.035) were associated with improved overall survival (Table 2). For the three features which were marginally significant for overall survival, tumors with high radial gradient border SD (Figure 1A, HR = 1.92; 95% CI 0.90-4.11, p = 0.092) and radial deviation tumor SD (Figure 1E, HR = 2.00; 95% CI 0.92-4.34, p = 0.078) were associated with poor overall survival while high radial gradient outside-tumor separation mean was associated with improved overall survival (Figure 1B, HR = 0.48; 95% CI 0.22-1.06, p = 0.068).

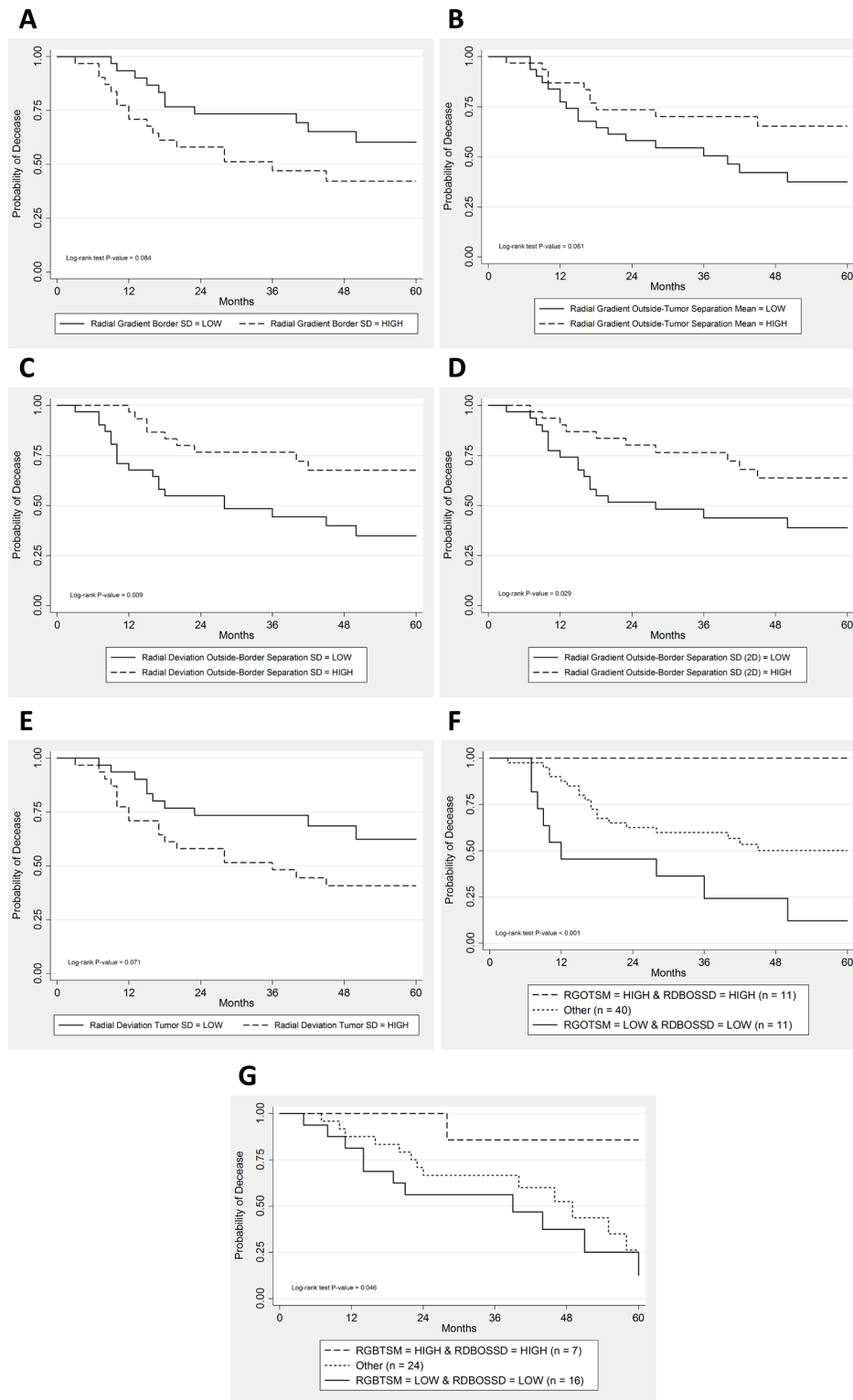
In an exploratory analysis, we found low correlation between the features that were previously published in these cohorts (entropy ratio and convexity) [2] to radial deviation and gradient features in this analysis (Pearson correlation coefficient < 0.35 for all features). Hence, the radial gradient and deviation features provide orthogonal information to previously identified features.

### Multivariable analyses

To reduce the number of image features to the most meaningful subset associated with overall survival, we applied a stepwise backward elimination model using a threshold of 0.01 to identify a parsimonious model. All five features and tumor volume were considered for the inclusion in the final model. The two features that remained in a feature-only multivariable model were radial deviation outside-border separation SD (HR = 0.25; 95% CI 0.11-0.58, p = 0.001) and radial gradient

outside-tumor separation mean (HR = 0.29; 95% CI 0.12-0.66, p = 0.003). To control for potential confounding, these two features were included in a multivariable Cox Model that included age, sex, and stage; and both features remained statistically significant (Table 2). We further analyzed these two features by demographics and imaging parameters and generated contingency tables. In the training cohort, none of the demographics or imaging parameters were significantly associated with radial deviation outside-border separation SD and radial gradient outside-tumor separation mean (Table 3).

To determine if these findings could be replicated in an external patient cohort, we analyzed these two features in a test cohort (Table 2) using the median threshold values obtained from the training cohort and found that radial deviation outside-border separation SD was statistically significant (Supplementary Figure 1B, HR = 0.36; 95% CI 0.16-0.81, p = 0.014) but radial gradient outside-tumor separation mean was not found to be statistically significant (Supplementary Figure 1A, HR = 0.75; 95% CI 0.28-2.03 p = 0.575) (Table 2). However, for both features, the point estimates were inversely associated with risk of death. When these two features were included in a multivariable Cox model that included age, sex and stage, radial deviation outside-border separation SD was statistically significant (HR = 0.40; 95% CI 0.17-0.97, p = 0.042) along with age (HR = 2.65; 95% CI 1.07-6.60, p = 0.035) and stage (HR = 3.35; 95% CI 1.34-8.36, p = 0.010) (Table 2). Additionally, among early stage lung cancer



**Figure 1: Kaplan-Meier survival curves for the following features. (A)** Radial gradient border standard deviation in the training cohort. **(B)** Radial gradient outside-tumor separation mean in the training cohort. **(C)** Radial deviation outside-border separation standard deviation in the training cohort. **(D)** Radial gradient outside-border separation standard deviation (2D) in the training cohort. **(E)** Radial deviation tumor standard deviation in the training cohort. **(F)** For the combination of radial gradient outside-tumor separation mean (RGOTSM) and radial deviation outside-border separation standard deviation (RDOBSSD) features in the training cohort. Hazard ratio with 95% confidence interval is calculated for the entire cohort (HR = 3.65; 95% CI (1.89–7.05)). **(G)** For the combination of radial gradient outside-tumor separation mean (RGOTSM) and radial deviation outside-border separation standard deviation (RDOBSSD) features in the test cohort.

**Table 3: Demographics and imaging parameters by image features in training cohort**

Covariate	Radial gradient outside-tumor separation mean			Radial deviation outside-border separation SD		
	LOW	HIGH	P- Value	LOW	HIGH	P- Value
<b>Sex, N (%)</b>						
Female	14 (45.2)	16 (53.3)	0.612	13 (43.3)	17 (54.8)	0.446
Male	17 (54.8)	14 (46.7)		17 (56.7)	14 (45.2)	
<b>Age, N (%)</b>						
< 65	8 (25.8)	12 (40.0)	0.283	8 (26.7)	12 (38.7)	0.416
≥ 65	23 (74.2)	18 (60.0)		22 (73.3)	19 (61.3)	
<b>Stage, N (%)</b>						
I/II	21 (67.8)	23 (77.4)	0.570	23 (77.4)	21 (67.7)	0.570
III/IV	10 (32.2)	7 (22.6)		7 (22.6)	10 (32.3)	
<b>5- year survival, %</b>	37.4%	65.3%	0.061	34.9%	67.7%	<b>0.009</b>
<b>Voltage, kVp, N (%)</b>						
120	28 (90.3)	29 (96.7)	0.612	26 (86.7)	31 (100.0)	0.053
130 or 140	3 (9.7)	1 (3.3)		4 (13.3)	0 (0)	
<b>Convolution kernel, N (%)</b>						
A,B	0 (0)	0 (0)	0.270	0 (0)	0 (0)	0.700
B30s,B60f,B70s	2 (6.7)	0(0)		1(3.2)	1(3.3)	
B30f	6 (20.0)	2 (6.4)		3 (9.7)	5 (16.7)	
B40f	7 (23.3)	12 (38.7)		12 (38.7)	7 (23.3)	
B41f	10 (33.3)	11 (35.5)		9 (29.0)	12 (40.0)	
Other	5 (16.7)	6 (19.4)		6 (19.4)	5 (16.7)	
<b>Interpolated slice thickness, N (%)</b>						
1.5 mm	0 (0)	0 (0)	0.189	0 (0)	0 (0)	0.861
2.0 mm	5 (16.1)	3 (10.0)		3 (10.0)	5 (16.1)	
2.5 mm	17 (54.8)	23 (76.7)		20 (66.7)	20 (64.5)	
3.0 mm	9 (29.1)	4 (13.3)		7 (23.3)	6 (19.4)	
<b>Pixel resolution, tertiles N (%)</b>						
< 0.6926 mm	7 (22.6)	13 (43.3)	0.146	7 (23.3)	13 (41.9)	0.172
≥ 0.6926 and < 0.7785 mm	10 (32.3)	10 (33.3)		13 (43.3)	7 (22.6)	
> 0.7785 mm	14 (45.1)	7 (23.4)		10 (33.4)	11 (35.5)	

<sup>1</sup>Numbers inside parenthesis are the percentage values.



**Table 4: Association between semantic features and radial gradient and radial deviation features**

Feature No.	Feature name	<i>Lobulation</i>			P- Value
		Absent	Present		
48	<b>Radial gradient outside-border separation SD (2D), N (%)</b>				
	LOW	27 (60.0)	4 (23.5)		<b>0.021</b>
	HIGH	18 (40.0)	13 (76.5)		
<i>Pleural attachment</i>					
		Absent	Present		
20	<b>Radial gradient border SD, N (%)</b>				
	LOW	28 (62.2)	3 (17.7)		<b>0.004</b>
	HIGH	17 (37.8)	14 (82.3)		
<i>Border definition</i>					
		Well defined	Poorly defined	Other <sup>1</sup>	
20	<b>Radial gradient border SD, N (%)</b>				
	LOW	13 (81.3)	8 (42.1)	10 (37.0)	<b>0.015</b>
	HIGH	3 (18.7)	11 (57.9)	17 (63.0)	
42	<b>Radial deviation outside-border separation SD<sup>2</sup>, N (%)</b>				
	LOW	4 (25.0)	9 (47.4)	18 (66.7)	<b>0.029</b>
	HIGH	12 (75.0)	10 (52.6)	9 (33.3)	
48	<b>Radial gradient outside-border separation SD (2D), N (%)</b>				
	LOW	4 (25.0)	14 (73.7)	13 (48.2)	<b>0.018</b>
	HIGH	12 (75.0)	5 (26.3)	14 (51.8)	

<sup>1</sup>Tumor with neither a well or poorly-defined border.

<sup>2</sup>This feature was replicated and found to be statistically significantly associated with survival in both the training cohort and test cohort.

patients (stage I and II), we found that radial deviation outside-border separation SD was statistically significantly associated with survival in the training cohort (log-rank p-value = 0.031) and marginally significant in the test cohort (log-rank p-value = 0.097). None of the patient demographics or imaging parameters were significantly associated with radial gradient outside-tumor separation mean and radial deviation outside-border separation SD except pixel resolution was found to be significantly associated with radial gradient outside-tumor separation mean in test cohort (p = 0.010, Supplementary Table 3).

### Combinatorial analyses

In exploratory analyses, we assessed the combinatorial effects of radial deviation outside-border separation SD and radial gradient outside-tumor separation in the training and test cohorts (Figure 1F and

1G, respectively and Supplementary Table 2). In both cohorts we found that patients who had high values (> median) for both features had statistically significantly better survival compared to patients who had low values (≤ median) for both features. We also explored the subset of early stage patients (stage I and II) and found that the combinatorial effect was also statistically significant for overall survival in the training cohort (Log-rank P-value = 0.020). Although the survival pattern was similar in the test cohort, it did not reach statistical significance (Log-rank P-value = 0.19).

### Associations with semantic radiological features

We found three RD/RG radiomic features that were statistically significantly associated with three semantic features (Table 4): lobulation, pleural attachment, and border definition. Importantly, the replicated feature

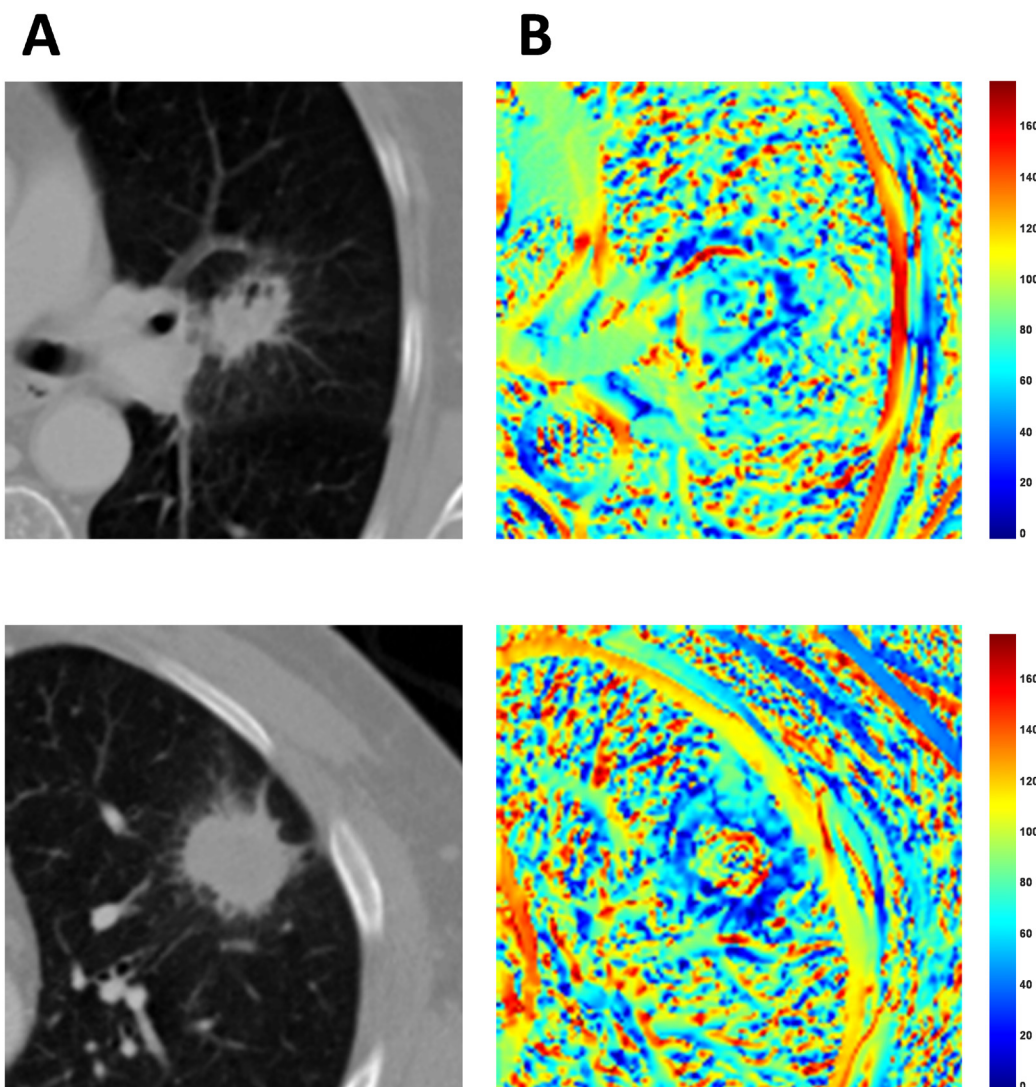
was significantly associated with border definition. Specifically, cancers with a well-defined border were significantly more likely to have high (> median) radial deviation outside-border separation SD. These analyses were restricted to the training cohort only.

## DISCUSSION

In this study we extracted radial gradient and radial deviation image features to determine whether they are associated with lung cancer patient survival. Of the 48 features that were extracted, 31 features were eliminated because they were not reproducible or they were redundant. The remaining 17 features were subjected to statistical analysis resulting in a parsimonious model containing two highly informative features associated with lung cancer survival. One of the two features (radial

deviation outside-border separation SD) was replicated and found to be statistically significantly associated with overall survival in a separate external cohort (test cohort) of lung cancer patients.

Radiomics is motivated by the premise that quantitative image features reflect the underlying pathophysiology of tumors. In Figure 2A and 2B we present the volume of interest (VOI) and corresponding radial deviation maps for two patients with substantially different clinical outcomes. The patient (Figure 2A) with short survival was deceased after 9 months and had a low (< median) radial deviation outside-border separation SD value while the second patient (Figure 2B) was still alive after 60 months had a high (>median) radial deviation outside-border separation SD value. In the original CT-image, both patients have similarly-sized tumors that are speculated; however, the VOI for each radial deviation



**Figure 2: Volume of interests (VOI) for two lung cancer patients with extreme differences in clinical outcomes.** Radial deviation image features for the corresponding VOIs for these lung cancer patients. The top row (A) is a tumor of a patient (Patient ID [PID]: 33) who deceased after 9 months and the second row (B) is a patient (PID: 75) with who had an ongoing survival after 60 months.

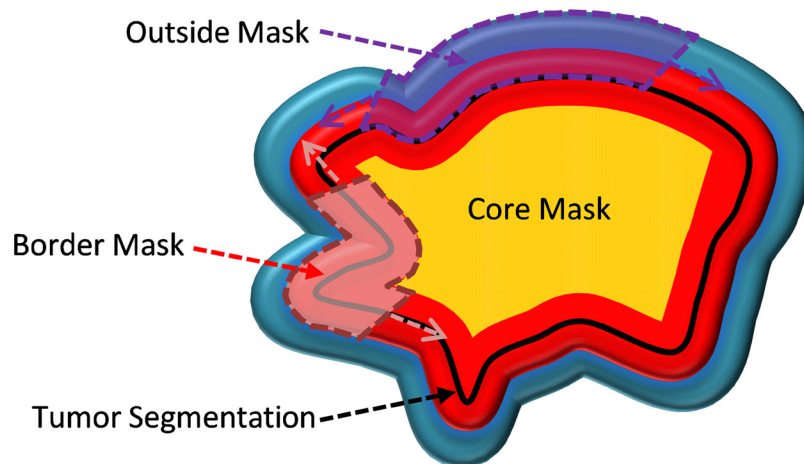


image have considerably different heat map appearances. By quantifying and analyzing these differences, as performed in this study, we have shown that RD/RG features may have clinical utility by differentiating patients with an aggressive disease and poor patient outcomes versus patients with more indolent disease and improved outcomes. Additionally, by analyzing the correlations of RD/RG features with semantic radiological features, we may have revealed their potential biological underpinnings. Specifically, we found three radial gradient and radial deviation features that were significantly associated with tumor lobulation, pleural attachment, and border definition (Table 4). The replicated feature, which was associated with lung cancer survival in both cohorts, was statistically significantly associated with border definition which has been previously reported to be a prognostic factor in lung cancer [11]. In the current analysis, patients who had well defined border definition were significantly associated with high radial deviation outside-border separation SD. As such, these analyses suggest that radial gradient and radial deviation features may be capturing clinically and biologically relevant radiological information of lung cancer tumors.

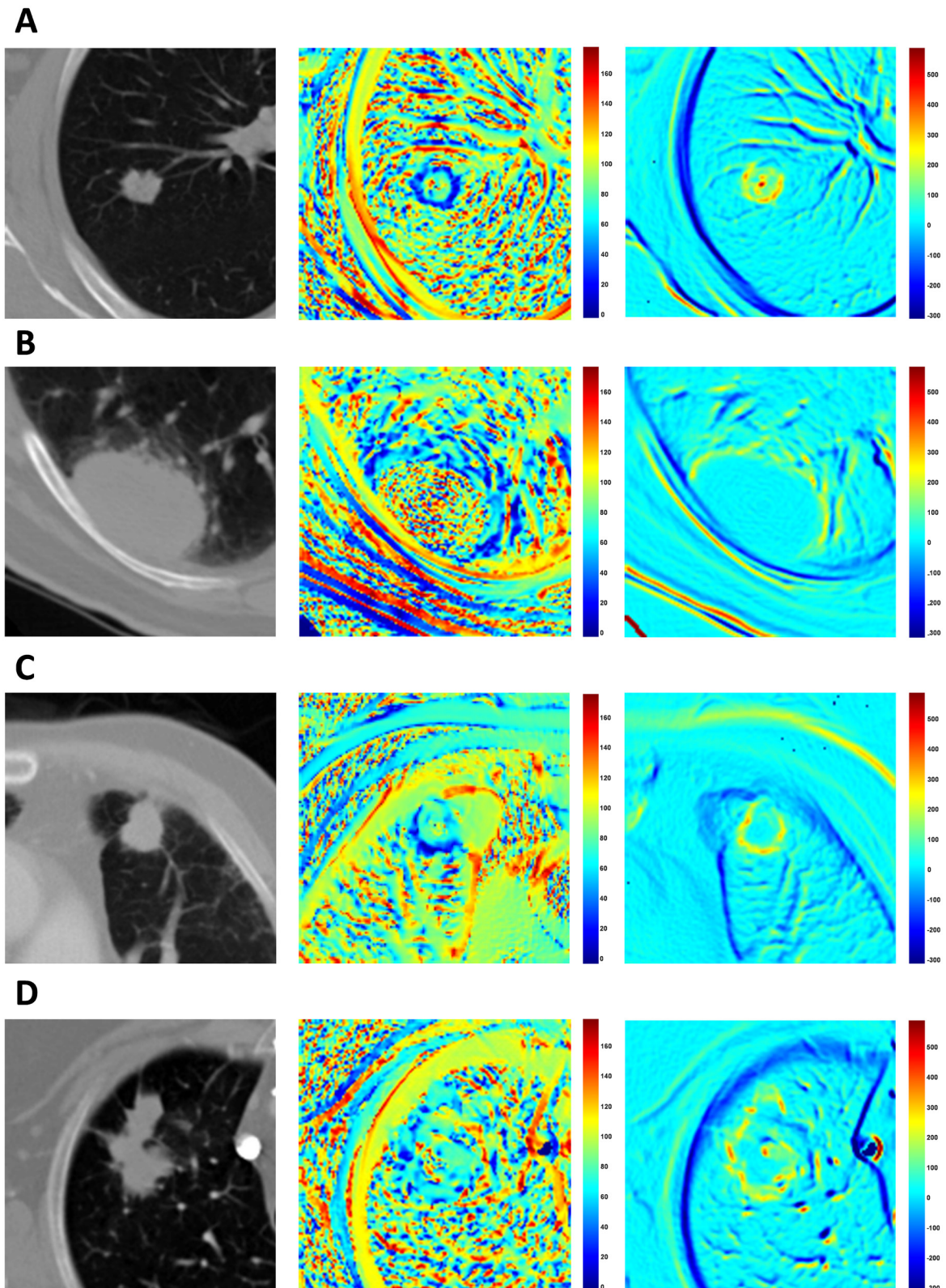
Radial image features have been previously applied in chest CT CAD systems to discriminate benign and malignant nodules, and optimize free parameters of tumor segmentation [19, 20]. However, in these previous studies, [19, 20] means and standard deviations were calculated from two different masks (region inside tumor and region outside tumor) on the radial gradient and radial deviation maps. By contrast, we calculated means and standard deviations from four different masks (tumor mask, border mask, core mask, and outside mask). To the best of our knowledge, the current study is the first to analyze RD/RG features for their association with lung cancer survival and their association with radiological semantic features.

In this study, we extracted and analyzed unique and new features from available training and test cohorts originally published by Grove et al. [2]. In the previous study, Grove et al. [2] reported that convexity and entropy ratio features were significantly associated with overall survival in the training cohort. Thus, we analyzed a model that included the convexity and entropy ratio features, RD/RG features, and patient characteristics and found that radial deviation outside-border separation SD (HR = 0.21), radial gradient outside-tumor separation mean (HR = 0.21), and entropy ratio (HR = 3.28) were statistically significantly associated with overall survival in the training cohort. However, when the remaining three features were analyzed in the test cohort, only radial deviation outside-border separation SD (HR = 0.34) was found to be statistically significant.

Quantitative image features have the potential to complement and improve current precision medicine. Limitations of tumor-based biomarkers are: they can be subjective to sampling bias due to the heterogeneous nature of tumors, the requirement of tumor specimens for biomarker testing, where the assays can be timely and expensive [2, 15, 21]. In contrast, radiomic features can be extracted in real-time from standard-of-care images, do not require timely and often expensive laboratory testing, are not subject to sampling bias and artifact, and are non-invasive. Importantly, radiomic analyses do not subject patients to additional radiation exposure since standard-of-care images are utilized, and radiomic features represent the phenotype of the entire tumor in 3D and not just the portion that was subjected to biomarker testing. Indeed, there is precedence that quantitative image features provide valuable and potentially translational information in lung cancer patient outcomes. Previous studies have shown that tumor shape and density are related to lung cancer survival [2, 13, 21, 22]. Additionally, as tumor



**Figure 3: Cartoon image of the four tumor masks.** The region inside the black line is the tumor mask, the yellow area is the core mask, the red area is the border mask, and the region outside the black line is the outside mask (combination of the half part of red region and whole blue region).



**Figure 4: Examples of radial deviation (middle column) and radial gradient (right column) maps. (A)** An example of a tumor which yields high contrast to the lung field and a round shape and hence has lower radial deviation angles pointing to the center of mass (PID: 150). **(B)** The standard deviation of radial deviation around border and outside regions are both high, making the separation value between them small (PID: 144). **(C)** The tumor shown has a round shape and has low radial deviation angle on border regions near the lung parenchyma but has heterogeneous values on the border to the lung wall (PID: 108). **(D)** Example tumor with an irregular shape making the radial deviation and radial gradient values heterogeneous (PID: 69).



shape becomes more eccentric, it has a higher probability of metastatic disease [2], and solid lesions that are differentiable from their outside environment and have high contrast edges and tend to be less aggressive [21]. Furthermore, tumors that are connected to lung wall are also associated with poor prognosis [21, 23].

We acknowledge some limitations and strengths to this study. First, we utilized an available set [2] of modestly sample sized training and test cohorts. However, we extracted and analyzed features that are unique from the prior work [2] and, importantly found an image feature to be significant in both the training and test cohorts. We applied a rigorous feature reduction approach to eliminate correlated and non-reproducible features, and we utilized a backward reduction approach to identify a single parsimonious model containing the most important features. We acknowledge in the combinatorial analyses that there were limited numbers of patients in the subgroups and we do not have disease-free survival data for these cohorts. A potential limitation of radiomic studies is the range of image acquisition parameters and modalities used [24] which can make it difficult to standardize image features and limit the robustness of computer-extracted features. In future studies, we will investigate interpolation methods to harmonize all data to a smaller range of slice thickness and pitch, which we hypothesize, will reduce some acquisition associated variability. However, we found that the image acquisition parameters were not associated with radial deviation outside-border separation SD for either cohort (Table 3 and Supplementary Table 3). Despite the fact that the study was limited to patients with lung adenocarcinoma, we removed potential histological differences in our analyses and we believe that this study had numerous strengths that outweigh the potential limitations.

In conclusion, this study identified a radial gradient and radial deviation image feature that was statistically significantly associated with lung cancer survival in both training and test cohorts even after adjusting for clinical covariates. Our analyses also revealed a novel combinatorial association of two features which differentiates patients with aggressive disease versus patients with indolent disease, and this was replicated in the test cohort. As such, these findings may have clinical utility to sub-stratify patients based on clinical outcome and identify patients that may need more aggressive treatment such as neo-adjuvant chemotherapy and aggressive follow-up and management. These features will require confirmation in additional studies and lung cancer patient cohorts.

## **MATERIALS AND METHODS**

### **Lung cancer patients**

This retrospective study was approved by the Institutional Review Boards at the University of South Florida and Maastricht University Medical Centre. There

were two separate cohorts used in this study that have been described elsewhere [2]. Briefly, the training cohort included 61 patients from the H. Lee Moffitt Cancer Center (MCC) & Research Institute, Tampa, Florida and the test cohort included 47 patients from the Maastricht Radiation Oncology Clinic (MAASTRO), Maastricht, Netherlands. All patients were diagnosed with lung adenocarcinoma and underwent surgical resection as first course of therapy. Pre-treatment contrast enhanced CT scans were acquired within two months prior to surgery. Both cohorts included diagnostic pre-treatment contrast-enhanced CT scans acquired between 2006 and 2009 and clinical data including demographics, histology, stage, and vital status information. Follow-up for vital status information occurs annually through passive and active methods.

### **Patient data**

For the training cohort, clinical data were obtained from Moffitt's Cancer Registry, which abstracts self-reported patient data and clinical information from patient medical records. Follow-up information for vital status occurs annually through passive and active methods. For this analysis, vital status was updated for the Moffitt patients since the previously published report [2]. Pathologic TNM staging was utilized when available and clinical stage was used if pathologic staging was unknown. Smoking status was categorized as ever smoker (current or former smoker) or never smoker. Similar data were abstracted and databased from MAASTRO for the test cohort patients.

### **Tumor segmentation**

All tumors were segmented using an in-house single-click ensemble segmentation algorithm on the Lung Tumor Analysis (LuTA) software program platform (Definiens Developer XD<sup>®</sup>, Munich, Germany) [25]. After applying the single click approach, the tumor delineations were inspected and edited if needed by a resident expert radiologist. The lung and tumor mask images obtained from LuTA software program were then imported into MATLAB<sup>®</sup> (Mathworks, Natick, MA) for image feature extraction as described below.

### **Radial gradient and radial deviation maps and features**

Development of the radial gradient and radial deviation image features has been previously described [18, 19]. In our study, after the tumors were segmented and center of mass of the tumor was automatically detected, 48 features were extracted from the radial gradient and radial deviation maps (Supplementary Table 4) bounded by different masks (described below) which were derived from tumor delineation masks using morphological

operations. Since there were variations in image acquisition parameters, we performed tri-linear interpolation by a factor of two for scans acquired with a slice thickness of  $\geq 4$  mm on the z-axis to create homogeneous spacing between scans. Additionally, pixels were interpolated tri-linearly in x and y directions to 2.50 mm x 2.50 mm.

The masks used were ‘tumor mask’, ‘border mask’, ‘core mask’, and ‘outside mask’. The tumor mask was the region that was delineated semi-automatically using Definiens Developer XD<sup>®</sup> software (Definiens, Inc., Cambridge, MA). The border mask is a “doughnut-shaped” region that was created by subtracting the two masks which are formed by a dilation operation followed by an erosion operation on the tumor mask. The region obtained after the erosion operation is the core mask. Structural elements radii used for dilation and erosion morphological operations were 7.5 mm and 12.5 mm for smaller tumors (major axis length (2D) <100 mm) and 10.0 mm and 15.0 mm for larger sized tumors (major axis length (2D)  $\geq 100$  mm). The outside mask was created by implementing dilation to the tumor mask followed by the subtraction of the tumor mask from the dilated region. The structural element used for the dilation morphological operation was 17.5 mm pixels for smaller tumors and 22.5 mm for larger sized tumors (Figure 3). All masks were additionally bounded to the lung parenchyma mask so that the VOIs did not exceed outside of the lung region.

The features were created using the four masks for both 2-dimensional and 3-dimensional. All 2-dimensional (2D) features were computed on the slice which included the center of mass of the segmentations. The 3-dimensional features were a natural extension of the 2D features and were calculated after the first and the last slices of the tumor segmentations were removed in order to reduce the partial volume effects. The separation features were calculated as the difference divided by the sum inside the analyzed masks (e.g., the radial deviation mean outside-border separation feature is the difference of outside radial deviation mean and the border radial deviation mean divided by the sum of the two).

Example radial gradient and radial deviation maps for the slices that contain the center of mass of tumors are presented in Figure 4. The radial deviation values along the border lines are close to  $0^\circ$  when the tumors have a more spherical shape (both vectors point the same direction) and they are higher when the tumor is irregularly shaped (Figure 4A and 4D). The radial gradient values are also affected by shape along with the gradient contrast that the tumors have in respect to their microenvironment (outside region). As a result, round shaped lesions tended to have lower radial deviation values while irregularly-shaped lesions have higher radial deviation values (Figure 3A and 3D). When the lesions are attached to the lung wall, despite the fact that tumor shape along the border is restricted to a rounder shape, the low

gradient on the intersection of the wall to the tumor creates deviating radial deviation values and diverges (Figure 4B and 4C). The magnitude of the gradient along the lung wall is close to zero (no gradient), as such radial gradient values are also lower along the lung wall. Radial gradient along the border is also affected by the shape of tumor, but it is furthermore influenced by how solid the tumor is in respect to its outside microenvironment. Solid tumors favor higher gradient values on border regions while semi-solid tumors have lower values (Figure 4A and 4D).

### **Elimination of redundant and non-reproducible features**

To eliminate the non-reproducible features we used the Reference Image Database to Evaluate Therapy Response (RIDER) dataset and calculated concordance correlation coefficient (CCC) between test-re-test scans. The RIDER dataset is a National Cancer Center (NCI) sponsored project for the guidance of integrating quantitative features across different institutions. The dataset is publicly available in National Biomedical Imaging Archive [26]. A total of 32 patients with unenhanced test-retest chest CT scans were acquired within 15 minutes of each other. The CCCs were calculated to quantify the reproducibility between consecutive scans for patients. Theoretically, CCC values range from -1 to 1, where 1 indicates a perfect correlation between two variables.

We calculated the CCCs for the 48 radial features and eliminated features that had a CCC < 0.80. As a result, 15 were features were dropped and the remaining 33 features were assessed for correlation using Pearson’s correlation coefficient. When two or more features resulted in a Pearson’s correlation coefficient greater than 0.80, we eliminated the feature(s) with the higher absolute column-wise correlation mean. Subsequently, 16 features were eliminated and the remaining 17 features enumerated in Supplementary Table 4 were subjected to statistical analysis.

### **Radiological semantic features**

Radiological semantic features for the training cohort were extracted by a clinical radiologist (YL with more than 7 years of experience) who was blinded to survival status and RD/RG status of the patients. Supplementary Table 5 contains the 13 radiological semantic features that were extracted from the CT scans of the subjects. Briefly, in terms of morphologic characteristics, the presence or absence of fissure attachment (defined as a tumor that attaches to the fissure; tumor’s margin is obscured by the margin), pleural attachment (defined as tumor attaches to the pleura other than fissure; tumor’s margin is obscured by the pleura), lobulation, concavity, air bronchogram [27], calcification, attachment to vessel, and pleural retraction

were assessed. We also evaluated the following features, which have been defined elsewhere: dominant attenuation pattern [27], shape, border definition, spiculation [28, 29], bubble-like lucency [28], cavitation [30].

## Statistical analyses

All statistical analyses were performed using Stata/MP 14.2 (StataCorp LP, College Station, TX). All image features were dichotomized at their median value. Differences in image features by demographic features and semantic features were tested using Fisher's exact test for categorical variables and Student's t-test for continuous variables. Survival analyses were performed using Cox Proportional Hazard Regression, Kaplan-Meier curves, and log-rank tests. All survival data were right-censored at 5 years. To reduce the number of features and covariates to a single parsimonious model, a backward elimination approach was applied. The features that were identified as the most informative in the training cohort were then tested in the test cohort.

## Abbreviations

RD: Radial Deviation; RG: Radial Gradient; SD: Standard Deviation; NSCLC: Non-small Cell Lung Cancer; CT: Computed Tomography; VOI: Volume of Interest; CAD<sub>c</sub>: Computer-aided Detection; CAD<sub>x</sub>: Computer-aided Diagnosis; HR: Hazard Ratio; CI: Confidence Interval; MCC: H. Lee Moffitt Cancer Center; MAASTRO: Maastricht Radiation Oncology Clinic; LuTA: Lung Tumor Analysis Software; 2D: 2-Dimensional; 3D: 3-Dimensional; CCC: Concordance Correlation Coefficient; RIDER: Reference Image Database to Evaluate Therapy Response; NCI: National Cancer Center.

## Author contributions

I.T. initialized the study, implemented the algorithms, performed the experiments, and co-wrote the manuscript. O.S. identified and generated the training cohort and critically reviewed the manuscript. A.G. supervised the implementation of the feature extraction algorithms and critically reviewed the manuscript. H.W. extracted images and edited delineations. Y.L. extracted radiological semantic features. Y.B. critically reviewed the manuscript. P.L. identified and generated the test cohort and critically reviewed the manuscript. R.J.G. obtained funding and critically reviewed the manuscript. M.B.S. conceptualized and supervised the research, obtained funding, analyzed the data, and co-wrote the manuscript.

## ACKNOWLEDGMENTS

None.

## CONFLICTS OF INTEREST

R.J.G. is a member of the Advisory Board at HealthMyne, Inc.

## FUNDING

Funding support came from the James & Esther King Biomedical Research Program-Team Science Project (2KT01 to Dr. Gillies); the National Cancer Institute (NCI) (U01-CA143062 to Dr. Gillies); and the NCI Early Detection Research Network (U01-CA200464 to Drs. Gillies and Schabath). This work has also been supported in part by a Cancer Center Support Grant (CCSG) at the H. Lee Moffitt Cancer Center and Research Institute; an NCI designated Comprehensive Cancer Center (grant number P30-CA76292).

## REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2017. *CA Cancer J Clin.* 2017; 67:7-30.
2. Grove O, Berglund AE, Schabath MB, Aerts HJ, Dekker A, Wang H, Velazquez ER, Lambin P, Gu Y, Balagurunathan Y, Eikman E, Gatenby RA, Eschrich S, Gillies RJ. Quantitative computed tomographic descriptors associate tumor shape complexity and intratumor heterogeneity with prognosis in lung adenocarcinoma. *PLoS One.* 2015; 10:e0118261.
3. Mirsadraee S, Oswal D, Alizadeh Y, Caulo A, van Beek E Jr. The 7th lung cancer TNM classification and staging system: Review of the changes and implications. *World J Radiol.* 2012; 4:128-134.
4. Pao W, Girard N. New driver mutations in non-small-cell lung cancer. *Lancet Oncol.* 2011; 12:175-180.
5. Schabath MB, Thompson ZJ, Gray JE. Temporal trends in demographics and overall survival of non-small-cell lung cancer patients at Moffitt Cancer Center from 1986 to 2008. *Cancer Control.* 2014; 21:51-56.
6. Kachroo S, Tong L, Spitz MR, Xing Y, Merriman K, Zhu DK, Fueger J, Amos CI, Etzel CJ. Trends in prevalence of prognostic factors and survival in lung cancer patients from 1985 to 2004 at a tertiary care center. *Cancer Detect Prev.* 2008; 32:101-108.
7. Dela Cruz CS, Tanoue LT, Matthay RA. Lung cancer: epidemiology, etiology, and prevention. *Clin Chest Med.* 2011; 32:605-644.
8. Birim O, Kappetein AP, van Klaveren RJ, Bogers AJ. Prognostic factors in non-small cell lung cancer surgery. *Eur J Surg Oncol.* 2006; 32:12-23.
9. Ries LAG. (2007). Cancer survival among adults : U.S. SEER program, 1988-2001, patient and tumor characteristics. (Bethesda, MD: U.S. Department of Health and Human Services, National Institutes of Health, National Cancer Institute).



10. Wang H, Schabath MB, Liu Y, Stringfield O, Balagurunathan Y, Heine JJ, Eschrich SA, Ye Z, Gillies RJ. Association between computed tomographic features and Kirsten rat sarcoma viral oncogene mutations in patients with stage I lung adenocarcinoma and their prognostic value. *Clin Lung Cancer*. 2016; 17:271-278.
11. Wang H, Schabath MB, Liu Y, Berglund AE, Bloom GC, Kim J, Stringfield O, Eikman EA, Klippenstein DL, Heine JJ, Eschrich SA, Ye Z, Gillies RJ. Semiquantitative computed tomography characteristics for lung adenocarcinoma and their association with lung cancer survival. *Clin Lung Cancer*. 2015; 16:e141-e163.
12. Coroller TP, Agrawal V, Narayan V, Hou Y, Grossmann P, Lee SW, Mak RH, Aerts HJ. Radiomic phenotype features predict pathological response in non-small cell lung cancer. *Radiother Oncol*. 2016; 119:480-486.
13. Ganeshan B, Panayiotou E, Burnand K, Dizdarevic S, Miles K. Tumour heterogeneity in non-small cell lung carcinoma assessed by CT texture analysis: a potential marker of survival. *Eur Radiol*. 2012; 22:796-802.
14. Aerts HJ, Grossmann P, Tan Y, Oxnard GG, Rizvi N, Schwartz LH, Zhao B. Defining a radiomic response phenotype: a pilot study using targeted therapy in NSCLC. *Sci Rep*. 2016; 6:33860.
15. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. 2016; 278:563-577.
16. Wang H, Schabath MB, Liu Y, Han Y, Li Q, Gillies RJ, Ye Z. Clinical and CT characteristics of surgically resected lung adenocarcinomas harboring ALK rearrangements or EGFR mutations. *Eur J Radiol*. 2016; 85:1934-1940.
17. Liu Y, Balagurunathan Y, Atwater T, Antic S, Li Q, Walker RC, Smith GT, Massion PP, Schabath MB, Gillies RJ. Radiological image traits predictive of cancer status in pulmonary nodules. *Clin Cancer Res*. 2017; 23:1442-1449.
18. Hardie RC, Rogers SK, Wilson T, Rogers A. Performance analysis of a new computer aided detection system for identifying lung nodules on chest radiographs. *Med Image Anal*. 2008; 12:240-258.
19. Messay T, Hardie RC, Rogers SK. A new computationally efficient CAD system for pulmonary nodule detection in CT imagery. *Med Image Anal*. 2010; 14:390-406.
20. Messay T, Hardie RC, Tuinstra TR. Segmentation of pulmonary nodules in computed tomography using a regression neural network approach and its application to the Lung Image Database Consortium and Image Database Resource Initiative dataset. *Med Image Anal*. 2015; 22:48-62.
21. Gevaert O, Xu JJ, Hoang CD, Leung AN, Xu Y, Quon A, Rubin DL, Napel S, Plevritis SK. Non-small cell lung cancer: identifying prognostic imaging biomarkers by leveraging public gene expression microarray data-methods and preliminary results. *Radiology*. 2012; 264:387-396.
22. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, Bussink J, Monshouwer R, Haibe-Kains B, Rietveld D, Hoebers F, Rietbergen MM, Leemans CR, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*. 2014; 5:4006.
23. Shultz DB, Trakul N, Abelson JA, Murphy JD, Maxim PG, Le QT, Loo BW Jr, Diehn M. Imaging features associated with disease progression after stereotactic ablative radiotherapy for stage I non-small-cell lung cancer. *Clin Lung Cancer*. 2014; 15:294-301.e3.
24. Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, Forster K, Aerts HJ, Dekker A, Fenstermacher D, Goldgof DB, Hall LO, Lambin P, et al. Radiomics: the process and the challenges. *Magn Reson Imaging*. 2012; 30:1234-1248.
25. Gu Y, Kumar V, Hall LO, Goldgof DB, Li CY, Korn R, Bendtsen C, Velazquez ER, Dekker A, Aerts H, Lambin P, Li X, Tian J, et al. Automated delineation of lung tumors from CT images using a single click ensemble segmentation approach. *Pattern Recognit*. 2013; 46:692-702.
26. Zhao B, James LP, Moskowitz CS, Guo P, Ginsberg MS, Lefkowitz RA, Qin Y, Riely GJ, Kris MG, Schwartz LH. Evaluating variability in tumor measurements from same-day repeat CT scans of patients with non-small cell lung cancer. *Radiology*. 2009; 252:263-272.
27. Hansell DM, Bankier AA, MacMahon H, McLoud TC, Muller NL, Remy J. Fleischner Society: glossary of terms for thoracic imaging. *Radiology*. 2008; 246:697-722.
28. Zwirowich CV, Vedal S, Miller RR, Muller NL. Solitary pulmonary nodule: high-resolution CT and radiologic-pathologic correlation. *Radiology*. 1991; 179:469-476.
29. Glynn C, Zakowski MF, Ginsberg MS. Are there imaging characteristics associated with epidermal growth factor receptor and KRAS mutations in patients with adenocarcinoma of the lung with bronchioloalveolar features? *J Thorac Oncol*. 2010; 5:344-348.
30. Onn A, Choe DH, Herbst RS, Correa AM, Munden RF, Truong MT, Vaporciyan AA, Isobe T, Gilcrease MZ, Marom EM. Tumor cavitation in stage I non-small cell lung cancer: epidermal growth factor receptor expression and prediction of poor outcome 1. *Radiology*. 2005; 237:342-347.