




Marginal radiomics features as imaging biomarkers for pathological invasion in lung adenocarcinoma

Hwan-ho Cho^{1,2} · Geewon Lee^{3,4} · Ho Yun Lee^{3,5}  · Hyunjin Park^{2,6}

Received: 1 August 2019 / Revised: 21 October 2019 / Accepted: 8 November 2019 / Published online: 21 January 2020
© European Society of Radiology 2020

Abstract

Objectives Lung adenocarcinomas which manifest as ground-glass nodules (GGNs) have different degrees of pathological invasion and differentiating among them is critical for treatment. Our goal was to evaluate the addition of marginal features to a baseline radiomics model on computed tomography (CT) images to predict the degree of pathologic invasiveness.

Methods We identified 236 patients from two cohorts (training, $n = 189$; validation, $n = 47$) who underwent surgery for GGNs. All GGNs were pathologically confirmed as adenocarcinoma in situ (AIS), minimally invasive adenocarcinoma (MIA), or invasive adenocarcinoma (IA). The regions of interest were semi-automatically annotated and 40 radiomics features were computed. We selected features using L1-norm regularization to build the baseline radiomics model. Additional marginal features were developed using the cumulative distribution function (CDF) of intratumoral intensities. An improved model was built combining the baseline model with CDF features. Three classifiers were tested for both models.

Results The baseline radiomics model included five features and resulted in an average area under the curve (AUC) of 0.8419 (training) and 0.9142 (validation) for the three classifiers. The second model, with the additional marginal features, resulted in AUCs of 0.8560 (training) and 0.9581 (validation). All three classifiers performed better with the added features. The support vector machine showed the most performance improvement (AUC improvement = 0.0790) and the best performance was achieved by the logistic classifier (validation AUC = 0.9825).

Conclusion Our novel marginal features, when combined with a baseline radiomics model, can help differentiate IA from AIS and MIA on preoperative CT scans.

Key Points

- Our novel marginal features could improve the existing radiomics model to predict the degree of pathologic invasiveness in lung adenocarcinoma.

Keywords Lung adenocarcinoma · Classification · Tumor microenvironment · Quantitative evaluation · Machine learning

Hwan-ho Cho and Geewon Lee contributed equally to this work.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00330-019-06581-2>) contains supplementary material, which is available to authorized users.

✉ Ho Yun Lee
hoyunlee96@gmail.com

✉ Hyunjin Park
hyunjinp@skku.edu

¹ Department of Electronic and Computer Engineering, Sungkyunkwan University, Suwon, Republic of Korea

² Center for Neuroscience Imaging Research, Institute for Basic Science, Suwon, Republic of Korea

³ Department of Radiology and Center for Imaging Science, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea

⁴ Department of Radiology and Medical Research Institute, Pusan National University Hospital, Pusan National University School of Medicine, Busan, Republic of Korea

⁵ Department of Health Sciences and Technology, Samsung Advanced Institute for Health Sciences & Technology (SAIHST), Sungkyunkwan University, Seoul, Republic of Korea

⁶ School of Electronic and Electrical Engineering, Sungkyunkwan University, Suwon, Republic of Korea

Abbreviations

AIS	Adenocarcinoma in situ
AUC	Area under the curve
CDF	Cumulative distribution function
CT	Computed tomography
DFS	Disease-free survival
GGNs	Ground-glass nodules
GLCM	Gray-level co-occurrence matrix
HU	Hounsfield unit
IA	Invasive adenocarcinoma
ISZM	Intensity size zone matrix
LASSO	Least absolute shrinkage and selection operator
MIA	Minimally invasive adenocarcinoma
MSE	Mean squared error
RF	Random forest
ROC	Receiver operator characteristic
ROI	Region of interest
SVM	Support vector machine
VNC	Virtual non-contrast-enhanced

Introduction

With respect to the degree of invasion, lung adenocarcinomas are pathologically classified as adenocarcinoma in situ (AIS), minimally invasive adenocarcinoma (MIA), or invasive adenocarcinoma (IA) according to the 2011 Multidisciplinary Classification of Lung Adenocarcinomas [1]. AIS and MIA are associated with 5-year disease-free survival (DFS) rates of nearly 100% after complete resection [2], compared with a rate of 74.6% in patients with stage 1 IA [3]. Furthermore, this survival rate decreases with the increasing invasive tumor component. Some surgeons have suggested that AIS and MIA are treatable by sublobar resection [4, 5] instead of lobectomy, which is the current gold standard for resection of early-stage lung cancers [6]. Therefore, it is important to distinguish IA from AIS and MIA in a timely manner because treatment options may vary according to preoperative radiologic diagnosis.

Many radiologists have attempted to differentiate AIS, MIA, and IA. In recent decades, qualitative measurements of size, shape, and border definition have been used to distinguish ground-glass nodules (GGNs) [7]. However, these qualitative measures are usually binarized (i.e., spiculated or non-spiculated), and there are concerns regarding inter- or intra-observer agreement. Moreover, despite differences in the pathologic invasive component, AIS, MIA, and lepidic-predominant IA are all usually observed as GGNs with or without a small solid portion. There is considerable overlap across the spectrum of lung adenocarcinomas, which makes radiologic interpretation challenging for stratifying GGNs [8].

Radiomics analysis can provide high-dimensional quantification of the tumor and has the potential to overcome

ambiguities related to visual assessment of GGNs [9]. A large number of quantitative imaging features that reflect morphologic, intensity, and textural properties can be extracted from medical imaging based on computational algorithms. New features have been actively developed to better quantify tumor characteristics in radiomics [10–13]. State-of-the-art radiomics analysis has been extended to cover peritumoral or tumor marginal properties to capture the microenvironment and invasion factors in a given tumor [14, 15]. The present study focused on marginal information based on the results of prior studies emphasizing the importance of semantic marginal features in lung adenocarcinoma [7]. We proposed reformulated tumor margin features based on the probability theory to enhance the radiomics approach.

This study evaluated the performance of tumor margin features to predict the degree of pathologic invasiveness in preoperative computed tomography (CT) scans. We built a baseline radiomics model and another model that combined additional marginal features with the baseline model. The classification performance of the two models was compared using multicenter data.

Materials and methods

Patients

Institutional review board approval (#SMC 2017-09-045) was obtained for this retrospective study, with the need for informed consent waived. Patients who underwent complete resection of GGNs between 2003 and 2013 were identified from a lung cancer surgical registry database of the Department of Thoracic Surgery at Samsung Medical Center (Seoul, Korea). Patients underwent surgery if the GGN was larger than 8 mm, persistent, and there was evidence of malignancy such as nodule growth or increasing solid portion according to the Fleischner Society Guidelines for GGNs [16]. We also considered patient preference for aggressive surgery over conservative observation. Among the patients, we excluded those with GGNs having solid components ≥ 5 mm in diameter because a solid component is usually considered IA. Overall, 189 patients with pure GGNs or part-solid nodules with limited solid components (< 5 mm) on preoperative CT scans were included as the training group in this study [17]. Our sample size of the cohort was determined based on the formula described by Fleiss et al [18] for unequal sample size analysis. The estimated total number of patients determined by the use of the chi-squared test for multiple proportions was 174 by using a power of 0.90, alpha of 0.05, and the above proportional settings. The study period was determined to have a sufficient number of samples. Consecutive patients from a lung cancer surgical registry database of the Department of Thoracic Surgery at our institute

were recruited. External validation of our study results was performed using an independent dataset of 47 patients from a different institution, Pusan National University Hospital (Pusan, Korea). Institutional review board approval (#E-2015084) was obtained and informed consent was waived. As in the training group, patients with pure GGNs or part-solid nodules with a limited solid component (< 5 mm) were included in the validation group. All GGNs were confirmed as AIS, MIA, or IA. All patients underwent staging and surgery according to the 7th edition of TNM for lung cancer staging published by the IASLC [19].

CT imaging

All chest CT scans were performed before surgery. For all patients in the training group, CT images were obtained with the following parameters: detector collimation, 1.25 or 0.625 mm; 120 kVp; 150–200 mA; and reconstruction interval 1–2.5 mm. All patients in the validation group underwent a CT examination using a multidetector CT system with similar parameters as the training group: 120 kVp; 150–200 mA; and section thickness range of 0.625–2.5 mm for axial images. The CT imaging parameters of both groups are shown in Supplementary Table 1. Image data were reconstructed with a soft-tissue algorithm for mediastinal window ranges and a bone algorithm for lung window images. Because the CT model has changed over the 11-year research period, we excluded low-quality CT images. All CT images used in this study were obtained from relatively high-quality 16-channel multidetector CT scanners.

Region of interest specification

All CT images in both the training and validation groups were displayed at standard mediastinal (window width, 400 Hounsfield units [HU]; window level, 20 HU) and lung (window width, 1500 HU; window level, –700 HU) window settings. On serial axial CT images displayed at a lung window setting, the region of interest (ROI) of the whole tumor was segmented by two chest radiologists using a semi-automated process [20]. ROIs were drawn on transverse CT scans at reconstruction intervals of 1–2.5 mm from the top to the bottom of the tumor, thus covering the whole tumor. The whole tumor margin including the ground-glass component was defined as the ROI.

Radiomics features

Radiomics features were computed using the largest representative ROI slice specified in the previous section. Feature computation was performed using the open source software PyRadiomics [21]. A total of 40 features were computed. Tumor area, mass, density, 19 histogram-based features, 16 gray-level co-occurrence matrix (GLCM)-based features,

and two intensity size zone matrix (ISZM)-based features were calculated for each ROI [22–27]. The histogram-based features quantify the properties of the intratumoral intensity distributions. The histogram-based features were computed from 128-bin histograms calculated over the intratumoral intensity range. The GLCM features quantify textural information and reflect intratumoral heterogeneity using a 2D histogram with 128 bins. A total of eight matrices corresponding to eight two-dimensional (2D) directions with an offset of one were computed and then averaged to yield a single matrix. The averaged matrix was used to compute the GLCM features. The ISZM features also quantify texture using blobs of similar intensity and differing size. We constructed a 128 × 256 matrix in which the first dimension was binned intensity and the second dimension was the size of the blobs. The size was not quantized and if a blob was larger than 256 voxels, it was considered to have a size of 256 voxels. We considered four neighbors to define the size of the blob. More details can be found in Supplementary Table 2.

Tumor margin features using the cumulative distribution function

Conventionally, the threshold applied to CT intensity has been used to assess the definition of tumor margins. A tumor region might be defined as a set of voxels with intensities above the threshold. A well-defined tumor is likely to exhibit abrupt changes in intensity at the margin. If we apply various threshold levels from low to high, we are likely to see abrupt changes in the tumor region around the specific threshold level. An ill-defined tumor is likely to have irregular changes in intensity in the margin; thus, we are likely to see gradual changes in the tumor region around various threshold levels. We modeled such changes in tumor region with respect to various thresholds using the cumulative distribution function (CDF) in the probability theory, denoted as $F(x)$, where x was the threshold level. The CDF was built from the intensity histogram and $F(x)$ denotes the portion of voxels below a threshold x . The CDF function is a non-decreasing function starting from zero and ranging to one. For a well-defined tumor, we are likely to observe the CDF curve staying relatively flat and then suddenly increasing as with increasing threshold. The variation in CDF slope would be large. For an ill-defined tumor, we are likely to observe the CDF curve increasing slowly with respect to an increasing threshold. The variation in CDF slope would be small. Figure 1 shows the CDF plots for well-defined and ill-defined tumors. Our model was two-dimensional (2D). We computed the CDF curve from the intensity histogram of a representative slice. The slope of the CDF was measured from the full range (minimum to maximum) of intensities within the ROI and we computed the mean, standard deviation (SD), skewness, and kurtosis of the CDF slope as the tumor margin features.

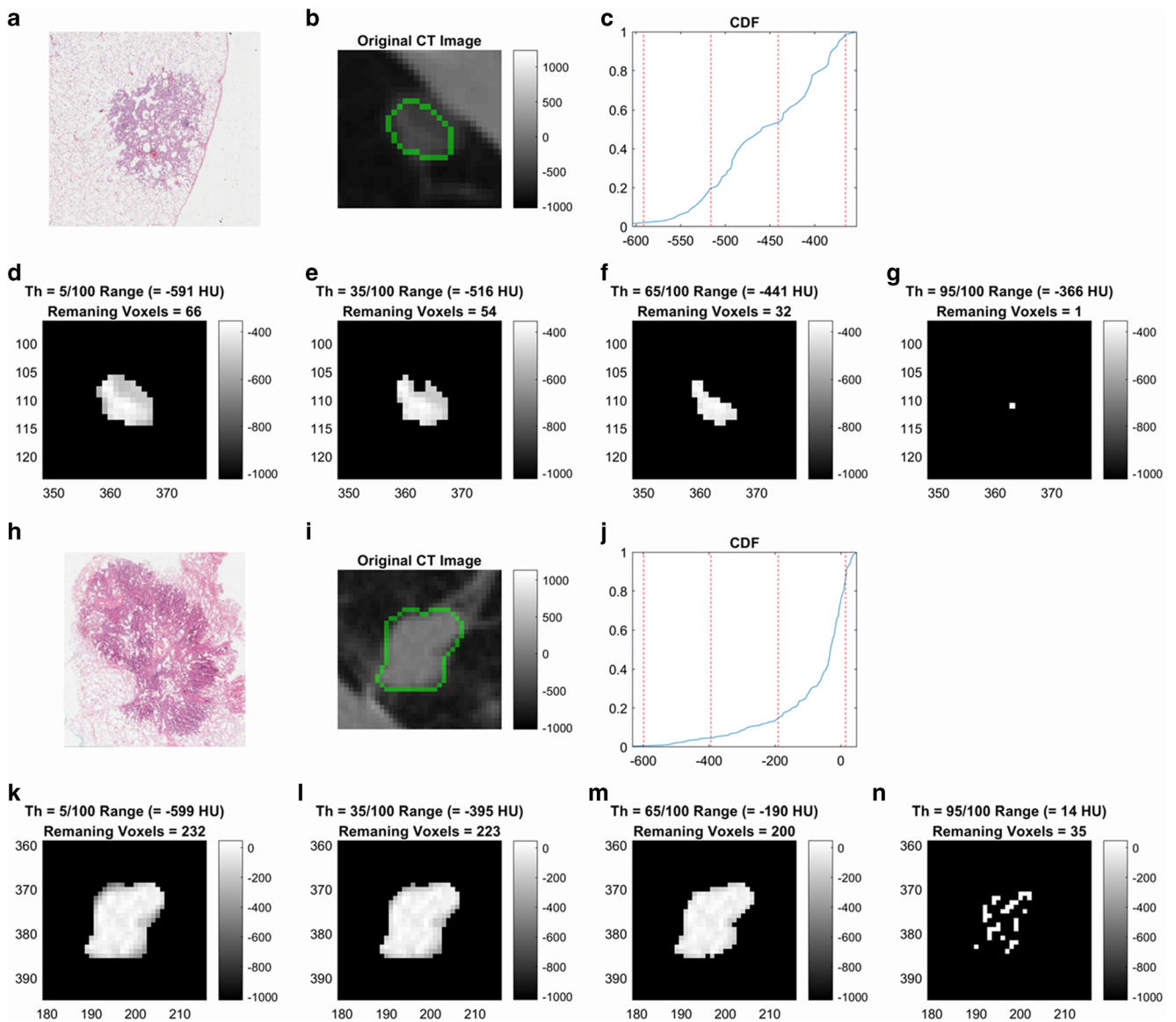


Fig. 1 Marginal features of the cumulative distribution function (CDF) for representative well- and ill-defined tumors. The first and last two rows illustrate ill- and well-defined tumors, respectively. Ill-defined tumor: **a** Histopathologic image of a 50-year-old female patient with an ill-defined GGN in the lingular division of the left upper lobe. Pathological diagnosis was confirmed as well-differentiated, lepidic-predominant, minimally invasive adenocarcinoma (MIA) with an invasive component of less than 2 mm. **b** A corresponding computed tomography (CT) image with the region of interest (ROI) overlaid in green. **c** The CDF plot in which the x-axis is in Hounsfield units (HU). Intratumor intensity range was divided into four intervals and the CDF plot has four vertical red lines that correspond to the 5/100, 35/100, 65/100, and 95/100 points of the intratumoral intensity range. **d–g** Correspond to applying the each

point of the range (= - 591, - 516, - 441, and - 366 HU) as the threshold, with gradually changing 66, 54, 32, and 1 voxels left after the threshold. Well-defined tumor: **h** Histopathologic image of a 73-year-old female patient with a well-defined GGN in the right upper lobe. Pathological diagnosis was confirmed as moderately differentiated, papillary and acinar pattern, invasive adenocarcinoma. **i** A corresponding CDF image with the ROI overlaid in green. **j** The CDF plot with the x-axis in HU. The CDF plot shows four vertical red lines that correspond to the 5/100, 35/100, 65/100, and 95/100 points of the intratumoral intensity range. **k–n** Correspond to applying each point of the range (= - 599, - 395, - 190, and 14 HU) as the threshold, with 232, 223, 200, and 35 voxels left after the threshold. With increasing threshold, the remaining voxels change abruptly after specific point (between - 190 HU and 14 HU in this case)

Reproducibility of ROIs and features

Two chest radiologists drew ROIs for all tumors. Cohen’s kappa was used to measure ROI repeatability and the reproducibility of imaging features was measured using intra-class correlation coefficients (ICCs).

Building a baseline model for predicting invasiveness

We performed feature selection from 40 radiomics features computed from the training cohort using a regression method combined with L1-norm regularization (i.e., least absolute shrinkage and selection operator [LASSO]) [28]. The

features' values were *z*-score-normalized and the response variable was set as to whether the tumor was IA or not (i.e., binary). For the optimization process, the lambda penalty was determined by applying a grid search, and the beta coefficient for each feature was decided using the gradient descent algorithm. For each lambda candidate, the mean squared error (MSE) was evaluated by tenfold cross-validation. We chose the lambda with the minimum MSE. The classification models included a logistic classifier, support vector machine (SVM), and random forest (RF). The inputs to the classifiers were the features selected by the LASSO. The logistic classifier was trained using multivariate logistic regression and the tumor class was decided by applying a threshold (0.5) to the logistic regression score. The SVM was trained with a second-order polynomial kernel. The RF was trained using tree-bagging with a 0.7 ratio of input data and 200 decision trees. For performance measurement, IA was defined as the positive class in the confusion matrix, and the accuracy, sensitivity, specificity, area under the curve (AUC) of the receiver operator characteristic (ROC) curve, adjusted *R*-squared value, and *p* value were evaluated.

Improving the baseline model by adding marginal features

To test whether the CDF-based marginal features could improve the baseline model, we added these features to those of the baseline model. The baseline model is referred to as model 1, while the improved model is referred to as model 2. The same training and performance evaluation procedures were used for both model 1 and model 2.

Applying the two models to the validation cohort

We applied the two models from the training data to an independent validation cohort. The same features from the training data were selected but the feature values were replaced with those computed from the validation cohort. The features values from the validation cohort were *z*-score-normalized using the mean and standard deviation values from the training data. The performance of the two models was evaluated with the same procedures described in the training data.

Statistical analysis

We used two-sample *t* tests to compare continuous-valued demographic information between training and validation cohorts. Chi-square tests were used to compare categorical variables between training and validation cohorts. To evaluate the statistical fitness of two proposed models, we adopted the *F*-test between the ground truth label and the diagnostic score for classification (e.g., logistic score of the logistic classifier). All

statistical evaluations were performed with Statistics and Machine Learning Toolbox in MATLAB (The MathWorks).

Results

Table 1 shows the demographic difference between the training and validation groups. Considering T staging, all patients in the training group were T1 status, whereas patients in the validation group were mostly T1 status (37 patients, 78.7%), with some T2 status (9 patients, 19.1%) or T3 status (1 patient, 2.1%). For N staging, only one patient in the training group (0.54%) was N1; all others were negative (N0) for lymph node status.

Feature selection identified five significant radiomics features. Table 2 shows the selected features and their corresponding beta coefficients, while Fig. 2 shows the MSE of the LASSO procedure. Three features (density, mass, and size zone variability) were also reported to be positively correlated with the extent of invasion on pathology in our previous study [29]. The classifier performance of the training and validation cohorts is shown in Table 3. Classification performance was improved by adding CDF features compared with the baseline model for all three classifiers in terms of AUC in the training cohort. The RF classifier yielded an AUC of 1 for both models, which could be due to overfitting. The same improvement in model 2 was observed in terms of AUC in the validation cohort. The highest performance improvement in the validation cohort was observed when we used the SVM classifier (i.e., AUC increase of 0.0790). The logistic classifier yielded the best performance (AUC 0.9825) in the validation cohort. On average, the AUC of the validation cohort was larger than that of the training cohort. This is partly due to the fact that the validation cohort had many extreme IA and AIS/MIA cases.

Regarding the reproducibility of ROIs and features, ROI repeatability in terms of Cohen's kappa was a mean (SD) of 0.8916 (0.0416), while the ICC of all features was a mean (SD) of 0.9311 (0.0760). The ICC of the nine features (i.e., five radiomics features and four CDF features) used in final modeling was a mean 0.9160 with SD 0.0617. (More details can be found in Supplementary Table 3.)

Discussion

Although lobectomy remains the standard treatment for lung cancers, there has been increasing evidence supporting limited surgery for lesions such as AIS and MIA. Nakayama et al [30] studied sublobar resections for 63 cT1N0M0 adenocarcinomas ≤ 2 cm in size. Overall survival was 95% for GGN and 69% for solid lesions, while recurrence-free survival was 100% in the former versus 57% in the latter. Another study by Fang et al [31], including 173 segmentectomy

Table 1 Demographic information of training and validation cohorts

Characteristics (%)	Training (n = 186)	Validation (n = 47)	p value
Sex			0.665
Male	85 (45.70%)	20 (42.55%)	
Female	101 (54.30%)	27 (57.45%)	
Age (years)	57.17 (9.50, 34–84)*	63.23 (8.84, 41–82)*	< 0.001
Tumor size (cm)			< 0.001
≤ 1	31 (16.67%)	5 (10.64%)	
1 < ≤ 2	108 (58.06%)	22 (46.81%)	
2 < ≤ 3	41 (22.04%)	13 (27.66%)	
> 3	6 (3.23%)	7 (14.89%)	
Histopathology			< 0.001
AIS	37 (19.89%)	3 (6.38%)	
MIA	61 (32.80%)	6 (12.77%)	
IA	88 (47.31%)	38 (80.85%)	

AIS adenocarcinoma in situ; MIA minimally invasive adenocarcinoma; IA invasive adenocarcinoma

*Data in parentheses indicate standard deviation and range

patients and 181 patients with wedge resection from three institutions, also showed similar results of GGN as an independent prognostic factor with no impact on survival according to the extent of resection. In other words, stratification of patients for limited surgery is a very important issue. However, there are substantial visual overlaps of imaging across the spectrum of lung adenocarcinomas, which makes it challenging to distinguish among GGNs. Therefore, this study proposed novel CDF-based marginal features based on the probability theory that reflected the degree of pathological invasion. The addition of marginal features enhanced the baseline radiomics approach and enabled differentiation of IA from AIS and MIA.

The baseline model included five features, many of which were previously reported to be correlated with degree of invasion [29]. The baseline model was improved by adding our CDF features. All three classifiers showed significant performance improvement in terms of AUC and accuracy. In the validation cohort, the SVM classifier showed the best performance improvement of 0.0790 (AUC). The best performance (AUC = 0.9825) was achieved by the logistic classifier.

Table 2 Selected features and beta coefficients of the baseline model

Selected feature	LASSO coefficient
Range	0.0611
GLCM entropy	− 0.0095
ISZM size zone variability	0.0305
Density	0.0177
Mass	0.0750

LASSO least absolute shrinkage and selection operator; GLCM gray-level co-occurrence matrix; ISZM intensity size zone matrix

Focusing on the tumor periphery, tumor cells, various stromal cells, extracellular matrix, and an extensive vascular network surrounding the tumor cells all make up the tumor microenvironment [32, 33]. This microenvironment is a dynamic area with continuous interactions between tumor cells and surrounding environment that plays a critical role in tumor metastasis and prognosis [32, 34]. Therefore, an understanding of the tumor microenvironment at the tumor periphery and its correlation with radiologic and pathologic features is essential. Although its underlying biology remains unclear, we investigated the influence of the tumor microenvironment by applying radiomics analysis and correlating the findings with pathologic invasiveness. The tumor microenvironment is

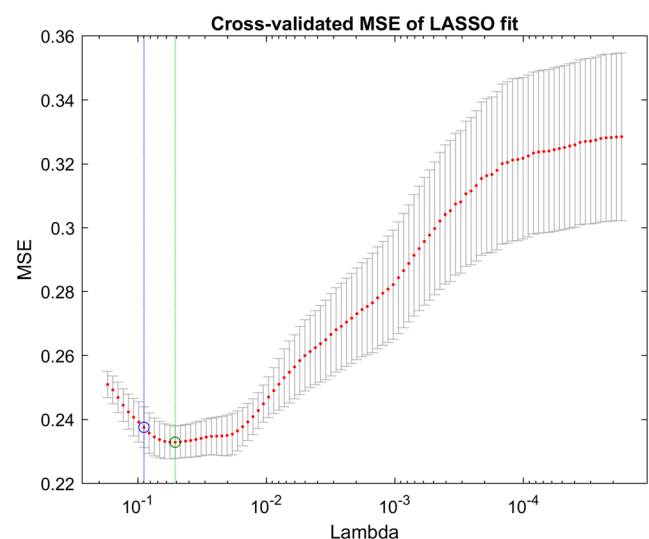


Fig. 2 MSE of the LASSO procedure. The green line indicates the minimum MSE point. Five significant features are selected at the minimum MSE point (Table 2)

Table 3 Classifier performance of the two models

		Accuracy	Sensitivity	Specificity	AUC	Adjusted <i>R</i> -squared	<i>p</i> value
Training Classifier							
Logistic							
	Model 1	0.6613	0.5909	0.7245	0.7490	0.1757	1.5725E-09
	Model 2	0.6828	0.6250	0.7347	0.7507	0.1781	1.1879E-09
SVM							
	Model 1	0.6935	0.7159	0.6735	0.7767	0.2201	4.8897E-11
	Model 2	0.7419	0.7841	0.7041	0.8291	0.3085	8.1370E-16
RF							
	Model 1	1.0000	1.0000	1.0000	1.0000	0.9326	6.3055E-110
	Model 2	1.0000	1.0000	1.0000	1.0000	0.9366	2.2911E-112
Validation Classifier							
Logistic							
	Model 1	0.8936	0.8947	0.8889	0.9766	0.2806	7.6634E-05
	Model 2	0.9149	0.8947	1.0000	0.9825	0.2700	1.0832E-04
SVM							
	Model 1	0.8085	0.7895	0.8889	0.8450	0.1907	3.5768E-03
	Model 2	0.8511	0.8421	0.8889	0.9240	0.3323	5.1925E-05
RF							
	Model 1	0.8298	0.8158	0.8889	0.9211	0.4264	3.9296E-07
	Model 2	0.8723	0.8684	0.8889	0.9678	0.6078	6.4906E-11

AUC area under the curve; *SVM* support vector machine; *RF* random forest

partly manifested in tumor margin information in terms of imaging. An ill-defined tumor is likely to have irregular changes in tumor margin as various threshold levels are tested. Our CDF-based feature is a useful approach for modeling such changes in the margin with respect to different threshold levels. Several features were computed from the CDF model, and we showed that these additional marginal features improved the baseline radiomics model for predicting the degree of pathological invasiveness.

Five radiomics features were selected in the baseline model. Our previous study identified the density, mass, and size zone variability of virtual non-contrast-enhanced (VNC) imaging as significant features to distinguish the degree of invasion of lung cancers [29]. We found that two additional features, the range and entropy of GLCM, were related to the degree of invasion. GLCM features are traditional texture measures that reflect spatial heterogeneity. GLCM textures have been identified as important features related to diagnosis, survival, and therapy response in many radiomics studies [35–37]. The entropy of GLCM measures the irregularity of intensity texture patterns. Regarding the histopathology, AIS and MIA are localized adenocarcinomas that exhibit a homogeneous lepidic growth pattern of tumor cells along the alveolar structures with an invasive component of less than 5 mm

[38]. In contrast, IAs harbor an invasive component measuring 5 mm or larger and are usually composed of multiple tumor subtypes of lepidic, acinar, papillary, solid, and micropapillary patterns. In other words, IAs might consist of different sub-compartments; thus, range and entropy of GLCM features could reflect such components of tumor compound [39, 40]. In addition, the high-intensity portion of the full intensity range is reportedly correlated with pathological microscopic invasion [41].

Recent radiomics studies have incorporated shape-based features into tumor characterization [9, 42]. However, many of them quantify the overall tumoral shape and do not focus exclusively on tumor margin. Recent studies reported that the quantification of the peritumoral microenvironment could lead to better modeling of the tumor [14, 15, 43, 44]. In a recent article, Beig et al suggested that densely packed tumor-infiltrating lymphocytes and tumor-associated stromal macrophages located at the margin of the tumor were associated with peritumoral radiomics features [45]. Our CDF feature was developed specifically for tumor margins and might be analogous to the internal thought process of human experts. An expert would apply this thought process to various threshold levels to assess tumor margins. Our CDF features were devised to mimic this process.

Inspired by the qualitative method to measure the degree of definement of a tumor margin in a clinical routine, our new features were designed to reflect the degree of definement in a quantitative manner. Our new features were mainly focused on the tumor margin, unlike the conventional radiomics features. Thus, the new features might be considered as complementary features to the existing conventional ones. We sought to add objective features reflecting the degree of definement. Our results showed the improved model with the new marginal features enhanced the prediction of invasiveness compared with the baseline model using conventional features that included morphological information. The baseline model performed well (AUC of 0.91 from validation) still, the added features improved the performance (AUC of 0.96 from validation). The gain in performance was rather incremental

(difference in AUC of 0.05). As implied in previous studies, we confirmed that the tumor margin contained important information, as shown with our new features, to explain the degree of invasion. In sum, the new marginal features might have a complementary and incremental impact to better predict the degree of invasion in lung adenocarcinoma.

The classifier performance was higher in the validation cohort compared with that in the training cohort. The inverted performance trend is unexpected but has also been shown in other radiomics studies [46, 47]. This inverted trend might be partly due to differences in demographic information between the two cohorts. As shown in Table 1, there was a significant difference in age, tumor size, and histopathology between the two groups. The validation cohort contained more cases of larger tumors (> 3 cm) and IA compared with those in the

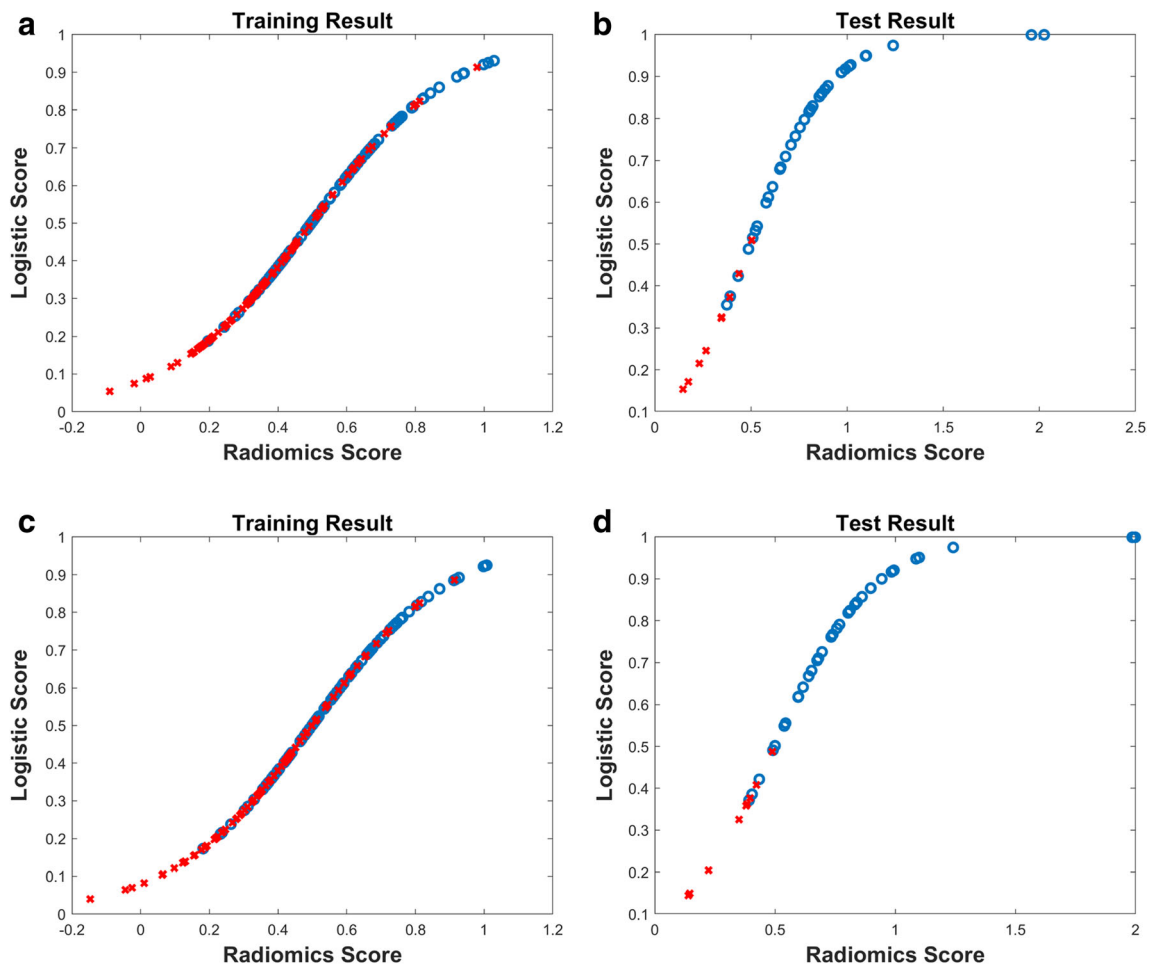


Fig. 3 Logistic regression classifier results from two cohorts. The red and blue dots indicate negative (MIA or AIS) and positive (IA) cases, respectively. **a, b** The classification results using the baseline model in two cohorts. **a** The results of the logistic regression classifier of the training cohort where blue and red dots overlap and, thus, lead to degraded classification performance. **b** From the test cohort, which shows less overlap between the blue and red dots and, thus, has better classification performance. **c, d** The classification results using the improved model. Similar

trends of more (**c**, training cohort) and less (**d**, test cohort) overlap are observed. There is an improvement from the baseline model (model 1) to the improved model (model 2), but the improvement is rather small (AUC 0.7490→0.7507 for the training cohort; 0.9766→0.9825 for the test cohort). Thus, the degree of overlap between colored dots is more difficult to observe when we compare regression plots in a column-wise fashion (i.e., **a, c** and **b, d**)

training cohort. The training cohort included many pathologically borderline cases of AIS and MIA, while the validation cohort had many pathologically extreme cases, as shown in Table 1. In other words, the validation group had more definitive cases of invasive adenocarcinoma. This makes classification easier, as there were fewer ambiguous samples. This was also confirmed on the logistic map of the two classes, as shown in Fig. 3. There was more overlap between the two classes in the training cohort compared with that in the validation cohort. Our results in the validation cohort could be inflated by bias in the validation cohort; thus, further validation is necessary. Another reason for the difference in the number of extreme cases may be the variability between the two institutions.

As the validation cohort was from a smaller hospital, it is possible that the surgeon from this hospital performed surgery only for GGNs that clearly showed invasiveness, while the surgeon from the larger hospital (training group) resected all GGNs regardless of invasiveness, thus including more AIS and MIA cases. In the same context, the validation cohort included tumors of larger size, in other words, more advanced tumors than were observed in the training cohort. This difference may reflect real-world medical practice because disease prevalence varies among different institutions.

Our study has several limitations. First, although ROIs were defined semi-automatically, intra- and inter-observer variability was possible. Development of an automatic method of ROI specification is planned in future research. Second, our sample size was relatively small; thus, further validation with a larger population is necessary. Third, our CDF features were two-dimensional models; a three-dimensional extension of the CDF features should be explored to determine if this model could better reflect tumor margins. Fourth, the training cohort included many pathologically borderline cases of AIS and MIA, while the validation cohort had many pathologically extreme cases. To minimize subjective variability, we used digital pathology. All tumor slides were scanned to produce a high-resolution digital image (0.25 lm/pixel at 40 \times) using the Aperio Slide Scanning System (ScanScope T3; Aperio Technologies Inc.). Two experienced lung pathologists jointly interpreted all tissue sections by virtual slides using the ImageScope viewing software (Aperio Technologies, Inc.) and a high-resolution monitor.

Our baseline radiomics model, which included range, GLCM entropy, ISZM size zone variability, density, and mass features, could distinguish IA from MIA and AIS. Furthermore, these reformulated marginal features improved the baseline radiomics model. Additional tumor margin features that reflect the degree of pathological invasion may contribute to accurate treatment planning.

Funding information This research was supported by Korea Health Industry Development Institute (HI17C0086), National Research

Foundation (NRF-2019R1H1A2079721 and NRF-2017M2A2A7A02018568), Ministry of Science and ICT (IITP-2019-2018-0-01798), IITP grant funded by the AI Graduate School Support Program (No. 2019-0-00421), and Institute for Basic Science (IBS-R015-D1).

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Ho Yun Lee

Conflict of interest The authors of this manuscript declare no relationships with any companies whose products or services may be related to the subject matter of the article.

Statistics and biometry One of the authors has significant statistical expertise.

Informed consent Written informed consent was waived by the Institutional Review Board.

Ethical approval Institutional Review Board approval was obtained.

Study subjects or cohort overlap Some study subjects have been previously reported in (Son JY, Lee HY, Kim JH, et al (2016) Quantitative CT analysis of pulmonary ground-glass opacity nodules for distinguishing invasive adenocarcinoma from non-invasive or minimally invasive adenocarcinoma: the added value of using iodine mapping. *Eur Radiol.* <https://doi.org/10.1007/s00330-015-3816-y>).

Methodology

- Retrospective
- Diagnostic or prognostic study
- Multicenter study

References

1. Austin JHM, Garg K, Aberle D et al (2013) Radiologic implications of the 2011 classification of adenocarcinoma of the lung. *Radiology.* <https://doi.org/10.1148/radiol.12120240>
2. Borczuk AC, Qian F, Kazeros A et al (2009) Invasive size is an independent predictor of survival in pulmonary adenocarcinoma. *Am J Surg Pathol.* <https://doi.org/10.1097/PAS.0b013e318190157c>
3. Zhang J, Wu J, Tan Q et al (2013) Why do pathological stage IA lung adenocarcinomas vary from prognosis?: a clinicopathologic study of 176 patients with pathological stage IA lung adenocarcinoma based on the IASLC/ATS/ERS classification. *J Thorac Oncol* 8:1196–1202. <https://doi.org/10.1097/JTO.0B013E31829F09A7>
4. Kates M, Swanson S, Wisnivesky JP (2011) Survival following lobectomy and limited resection for the treatment of stage I nonsmall cell lung cancer \leq 1 cm in size: a review of SEER data. *Chest* 139:491–496. <https://doi.org/10.1378/CHEST.09-2547>
5. Wisnivesky JP, Henschke CI, Swanson S et al (2010) Limited resection for the treatment of patients with stage IA lung cancer. *Ann Surg.* <https://doi.org/10.1097/SLA.0b013e3181c0e5f3>
6. MacMahon H, Naidich DP, Goo JM et al (2017) Guidelines for management of incidental pulmonary nodules detected on CT images: from the Fleischner Society 2017. *Radiology.* <https://doi.org/10.1148/radiol.2017161659>
7. Zhang Y, Shen Y, Qiang JW, Ye JD, Zhang J, Zhao RY (2016) HRCT features distinguishing pre-invasive from invasive

- pulmonary adenocarcinomas appearing as ground-glass nodules. *Eur Radiol* 26:2921–2928. <https://doi.org/10.1007/s00330-015-4131-3>
8. Kim HY, Shim YM, Lee KS, Han J, Yi CA, Kim YK (2007) Persistent pulmonary nodular ground-glass opacity at thin-section CT: histopathologic comparisons. *Radiology* 245:267–275. <https://doi.org/10.1148/radiol.2451061682>
 9. Yip SS, Aerts HJ (2016) Applications and limitations of radiomics. *Phys Med Biol* 61:R150–R166. <https://doi.org/10.1088/0031-9155/61/13/R150>
 10. Prasanna P, Tiwari P, Madabhushi A (2016) Co-occurrence of local anisotropic gradient orientations (CoLIAGe): a new radiomics descriptor. *Sci Rep* 6:37241. <https://doi.org/10.1038/srep37241>
 11. Ismail M, Hill V, Stasevych V et al (2018) Shape features of the lesion habitat to differentiate brain tumor progression from pseudoprogression on routine multiparametric MRI: a multisite study. *AJNR Am J Neuroradiol* 39:2187–2193. <https://doi.org/10.3174/ajnr.A5858>
 12. Grélard F, Baldacci F, Vialard A, Domenger J-P (2017) New methods for the geometrical analysis of tubular organs. *Med Image Anal* 42:89–101. <https://doi.org/10.1016/J.MEDIA.2017.07.008>
 13. Alilou M, Orooji M, Beig N et al (2018) Quantitative vessel tortuosity: a potential CT imaging biomarker for distinguishing lung granulomas from adenocarcinomas. *Sci Rep* 8:15290. <https://doi.org/10.1038/s41598-018-33473-0>
 14. Braman NM, Etesami M, Prasanna P et al (2017) Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI. *Breast Cancer Res* 19:57. <https://doi.org/10.1186/s13058-017-0846-1>
 15. Prasanna P, Patel J, Partovi S et al (2016) Radiomic features from the peritumoral brain parenchyma on treatment-naive multi-parametric MR imaging predict long versus short-term survival in glioblastoma multiforme: preliminary findings. *Eur Radiol*. <https://doi.org/10.1007/s00330-016-4637-3>
 16. Naidich DP, Bankier AA, MacMahon H et al (2013) Recommendations for the management of subsolid pulmonary nodules detected at CT: a statement from the Fleischner Society. *Radiology* 266:304–317. <https://doi.org/10.1148/radiol.12120628>
 17. Lee SM, Park CM, Song YS et al (2017) CT assessment-based direct surgical resection of part-solid nodules with solid component larger than 5 mm without preoperative biopsy: experience at a single tertiary hospital. *Eur Radiol* 27:5119–5126. <https://doi.org/10.1007/s00330-017-4917-6>
 18. Fleiss JL, Levin B, Paik MC (2013) Statistical methods for rates and proportions, 3rd edn. Wiley, Hoboken
 19. UyBico SJ, Wu CC, Suh RD et al (2010) Lung cancer staging essentials: the new TNM staging system and potential imaging pitfalls. *Radiographics* 30:1163–1181. <https://doi.org/10.1148/rg.305095166>
 20. Lee G, Park H, Sohn I et al (2018) Comprehensive computed tomography radiomics analysis of lung adenocarcinoma for prognostication. *Oncologist* 23:806–813. <https://doi.org/10.1634/theoncologist.2017-0538>
 21. Van Griethuysen JJM, Fedorov A, Parmar C et al (2017) Computational radiomics system to decode the radiographic phenotype. *Cancer Res* 77:e104–e107. <https://doi.org/10.1158/0008-5472.CAN-17-0339>
 22. Haralick RM, Shanmugam K, Dinstein I (1973) Textural features for image classification. *IEEE Trans Syst Man Cybern SMC-3*: 610–621. <https://doi.org/10.1109/TSMC.1973.4309314>
 23. Grove O, Berglund AE, Schabath MB et al (2015) Quantitative computed tomographic descriptors associate tumor shape complexity and intratumor heterogeneity with prognosis in lung adenocarcinoma. *PLoS One* 10:1–14. <https://doi.org/10.1371/journal.pone.0118261>
 24. Tixier F, Le Rest CC, Hatt M et al (2011) Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET/CT predicts response to concomitant radio chemotherapy in esophageal cancer. *J Nucl Med* 52:369–378. <https://doi.org/10.2967/jnumed.110.082404>
 25. Davnall F, Yip CSP, Ljungqvist G et al (2012) Assessment of tumor heterogeneity: an emerging imaging tool for clinical practice? *Insights Imaging* 3:573–589. <https://doi.org/10.1007/s13244-012-0196-6>
 26. de Hoop B, Gietema H, van de Vorst S et al (2010) Pulmonary ground-glass nodules: increase in mass as an early indicator of growth. *Radiology* 255:199–206. <https://doi.org/10.1148/radiol.09090571>
 27. Lee HY, Jeong JY, Lee KS et al (2012) Solitary pulmonary nodular lung adenocarcinoma: correlation of histopathologic scoring and patient survival with imaging biomarkers. *Radiology* 264:884–893. <https://doi.org/10.1148/radiol.12111793>
 28. Tibshirani R (1996) Regression selection and shrinkage via the Lasso. *J R Stat Soc B* 58:267–288
 29. Son JY, Lee HY, Kim JH et al (2016) Quantitative CT analysis of pulmonary ground-glass opacity nodules for distinguishing invasive adenocarcinoma from non-invasive or minimally invasive adenocarcinoma: the added value of using iodine mapping. *Eur Radiol*. <https://doi.org/10.1007/s00330-015-3816-y>
 30. Nakayama H, Yamada K, Saito H et al (2007) Sublobar resection for patients with peripheral small adenocarcinomas of the lung: surgical outcome is associated with features on computed tomographic imaging. *Ann Thorac Surg* 84:1675–1679. <https://doi.org/10.1016/J.ATHORACSUR.2007.03.015>
 31. Fang W XY, Zhong C, Chen Q (2014) The IASLC/ATS/ERS classification of lung adenocarcinoma—a surgical point of view. *J Thorac Dis* 6:S552–S560. <https://doi.org/10.3978/j.issn.2072-1439.2014.06.09>
 32. Wu T, Dai Y (2017) Tumor microenvironment and therapeutic response. *Cancer Lett* 387:61–68. <https://doi.org/10.1016/J.CANLET.2016.01.043>
 33. Hanahan D, Coussens LM (2012) Accessories to the crime: functions of cells recruited to the tumor microenvironment. *Cancer Cell* 21:309–322. <https://doi.org/10.1016/J.CCR.2012.02.022>
 34. Quail DF, Joyce JA (2013) Microenvironmental regulation of tumor progression and metastasis. *Nat Med* 19:1423–1437. <https://doi.org/10.1038/nm.3394>
 35. Aerts HJWL, Velazquez ER, Leijenaar RTH et al (2014) Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 5. <https://doi.org/10.1038/ncomms5006>
 36. Alcaide-Leon P, Dufort P, Geraldo AF et al (2017) Differentiation of enhancing glioma and primary central nervous system lymphoma by texture-based machine learning. *AJNR Am J Neuroradiol*. <https://doi.org/10.3174/ajnr.A5173>
 37. Kunimatsu A, Kunimatsu N, Kamiya K et al (2018) Comparison between glioblastoma and primary central nervous system lymphoma using MR image-based texture analysis. *Magn Reson Med Sci*. <https://doi.org/10.2463/mrms.mp.2017-0044>
 38. Travis WD, Brambilla E, Noguchi M et al (2011) International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society International Multidisciplinary Classification of Lung Adenocarcinoma. *J Thorac Oncol* 6:244–285. <https://doi.org/10.1097/JTO.0B013E318206A221>
 39. Gupta R, Phan CM, Leidecker C et al (2010) Evaluation of dual energy CT for differentiating intracerebral hemorrhage from iodinated contrast material staining. *Radiology*. <https://doi.org/10.1148/radiol.10091806>

40. Chae EJ, Song J-W, Seo JB et al (2008) Clinical utility of dualenergy CT in the evaluation of solitary pulmonary nodules: initial experience. *Radiology*. <https://doi.org/10.1148/radiol.2492071956>
41. Nomori H, Ohtsuka T, Naruke T, Suemasu K (2003) Histogram analysis of computed tomography numbers of clinical T1 N0 M0 lung adenocarcinoma, with special reference to lymph node metastasis and tumor invasiveness. *J Thorac Cardiovasc Surg* 126:1584–1589. [https://doi.org/10.1016/S0022-5223\(03\)00885-7](https://doi.org/10.1016/S0022-5223(03)00885-7)
42. Lee G, Lee HY, Park H 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. <https://doi.org/10.1016/j.ejrad.2016.09.005>
43. Lee AK, DeLellis RA, Silverman ML et al (1990) Prognostic significance of peritumoral lymphatic and blood vessel invasion in node-negative carcinoma of the breast. *J Clin Oncol* 8:1457–1465. <https://doi.org/10.1200/JCO.1990.8.9.1457>
44. Uematsu T (2015) Focal breast edema associated with malignancy on T2-weighted images of breast MRI: peritumoral edema, prepectoral edema, and subcutaneous edema. *Breast Cancer* 22: 66–70. <https://doi.org/10.1007/s12282-014-0572-9>
45. Beig N, Khorrami M, Alilou M et al (2019) Perinodular and intranodular radiomic features on lung CT images distinguish adenocarcinomas from granulomas. *Radiology* 290:783–792. <https://doi.org/10.1148/radiol.2018180910>
46. Beig N, Patel J, Prasanna P et al (2018) Radiogenomic analysis of hypoxia pathway is predictive of overall survival in Glioblastoma. *Sci Rep* 8:1–11. <https://doi.org/10.1038/s41598-017-18310-0>
47. Fan L, Fang M, Li Z et al (2019) Radiomics signature: a biomarker for the preoperative discrimination of lung invasive adenocarcinoma manifesting as a ground-glass nodule. *Eur Radiol* 29:889–897. <https://doi.org/10.1007/s00330-018-5530-z>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.