



A CT-based logistic regression model to predict spread through air space in lung adenocarcinoma

Chuanjun Li¹, Changsi Jiang², Jingshan Gong², Xiaotao Wu¹, Yan Luo², Guopin Sun¹

¹Department of Radiology, Pingshan District People's Hospital of Shenzhen, Shenzhen, China; ²Department of Radiology, Shenzhen People's Hospital (The Second Clinical Medical College, Jinan University; The First Affiliated Hospital, Southern University of Science and Technology), Shenzhen, China

Correspondence to: Jingshan Gong. Department of Radiology, Shenzhen People's Hospital (The Second Clinical Medical College, Jinan University; The First Affiliated Hospital, Southern University of Science and Technology), Shenzhen 518020, China. Email: jshgong@sina.com.

Background: Spread through air space (STAS) is a novel invasive pattern of lung adenocarcinoma and is also a risk factor for recurrence and worse prognosis of lung adenocarcinoma. This study aimed to develop and validate a computed tomography (CT)-based logistic regression model to predict STAS in lung adenocarcinoma.

Methods: This retrospective study was approved by the institutional review board of two centers and included 578 patients (462 from center I and 116 from center II) with pathologically confirmed lung adenocarcinoma. STAS was identified from 90 center I patients (19.5%) and 28 center II patients (24.1%) from. The maximum diameter, nodule area, and area of solid components in part-solid nodules were measured. Twenty-one semantic characteristics were assessed. Univariate analysis was used to select CT characteristics, which were associated with STAS in the patient cohort of center I. Multivariable logistic regression was used to develop a CT characteristics-based model on those variables with statistical significance. The model was validated in the validation cohort and then tested in the external test cohort (patients from center II). The diagnostic performance of the model was measured by area under the curve (AUC) of receiver operating characteristic (ROC).

Results: At univariate analysis, age and 11 CT characteristics, including the maximum diameter of the tumor, the maximum area of the tumor, the area and ratio of the solid component, nodule type, pleural thickening, pleural retraction, mediastinal lymph node enlargement, vascular cluster sign, and lobulation, specula were found to be significantly associated with STAS. The optimal logistic regression model included age, maximum diameter and ratio of solid component with odds ratio (OR) value of 0.967 (95% CI: 0.944–0.988), 1.027 (95% CI: 1.008–1.046) and 5.14 (95% CI: 2.180–13.321), respectively. This model achieved an AUC of 0.801 (95% CI: 0.709–0.892) and 0.692 (95% CI: 0.518–0.866) in the validation cohort and the external test cohort, respectively. The difference was not statistically significant ($P=0.280$).

Conclusions: CT-based logistic regression machine learning model could preoperatively predict STAS in lung adenocarcinoma with excellent diagnosis performance, which could be supplementary to routine CT interpretation.

Keywords: Spread through air space (STAS); lung adenocarcinoma; computed tomography (CT)

Submitted Jun 02, 2020. Accepted for publication Jul 16, 2020.

doi: 10.21037/qims-20-724

View this article at: <http://dx.doi.org/10.21037/qims-20-724>

Introduction

Spread through air space (STAS) is defined as a detachment of micropapillary clusters, solid nests, or single cells beyond the edge of the tumor into the air spaces in surrounding lung parenchyma (1). This detachment is considered a novel invasion pattern of lung adenocarcinoma after or beyond the infiltration of the myofibroblast stroma, lymph vascular, and pleura (2,3). Onozato and colleagues first reported that the tumor island is detached to a collection of tumor cells that can be observed within the alveolar space from the primary tumor mass by a distance of at least a few alveoli (4). After this phenomenon was verified by two substantial studies and correlated with recurrence-free survival (not related to the stage), STAS was introduced into the 2015 World Health Organization (WHO) classification (5). Even though STAS is a worse prognosticator, if patients with STAS-positive tumors undergo lobectomy resection instead of limited local resection, patients' prognosis can be significantly improved (6-8). Therefore, knowledge of STAS status before operation can facilitate surgeons in choosing an appropriate operation pattern for patients. Unfortunately, STAS is a histopathological finding, which can only be obtained after the operation. Recently, several reports have shown that several computed tomography (CT) characteristics were associated with STAS status (9-11). In one study, de Margerie-Mellon *et al.* studied 40 STAS-positive nodules and 40 STAS-negative nodules. From them, 203 subsolid nodules were identified; they found that the nodule diameter and the direct and relative diameters of the solid component were positively associated with STAS (9). In another study, Toyokawa *et al.* not only showed that a larger radiologic tumor diameter, vascular convergence, notch, pleural indentation, speculation, and the absence of ground-glass opacity (GGO) were associated with STAS in univariable analysis. They also found that the notch and the absence of GGO were risk factors of STAS with a combined odds ratio (OR) of 5.01 (10). Kim *et al.* analyzed CT features of 92 STAS-positive nodules and 184 STAS-negative nodules and found that the percentage of solid components was an independent predictor of STAS, and could obtain a sensitivity of 89.2% and a specificity of 60.3% using a cutoff value of 90% (11). Previously, we developed a CT-based radiomics machine learning model to predict STAS (12), and the model achieved an AUC of 0.754 (a sensitivity of 0.880 and a specificity of 0.588) for predicting STAS. This model showed that the CT-based radiomics could preoperatively predict STAS in lung adenocarcinoma with excellent diagnosis performance.

Therefore, this study aimed to evaluate the value of the CT characteristics logistic regression model for the preoperative assessment of lung adenocarcinoma STAS status.

Methods

Patients

This retrospective study was approved by the institutional review board of both centers, and the requirement for informed consent was waived. From April 2015 to April 2019, information from 695 consecutive patients (531 from the center I and 164 from center II) with surgical histopathologically confirmed lung adenocarcinoma were located on the two centers' electronic databases. We excluded patients who underwent CT examinations 3 months before the operation (n=19), received preoperative neoadjuvant chemotherapy (n=27) or preoperative biopsy with the theoretical possibility of infiltration or contamination through needles (n=32), or whose pathological sections were unsuitable for STAS detection (n=24), or who had more than one pathological confirmed tumor resected (n=15). The final study cohort included 578 patients (462 from center I and 116 from center II). *Figure 1* shows the patient cohort's workflow.

Histopathologic evaluation

For this study, two experienced pathologists who were blinded to the patient's clinical outcomes were asked to review the hematoxylin and eosin (HE) tissue section again according to the WHO definitions of STAS in consensus to determine the STAS status. STAS positive is defined as the discovery of tumor cells in the lung air spaces beyond the edge of the primary tumor. It mainly consists of the following three forms: (I) an air space filled by micro-nipple structure without central fibrovascular cores; (II) a solid nest, with air spaces filled by the solid component of the tumor; (III) air spaces filled by multiple discrete and discontinuous single cells (5).

CT image acquisition and assessment

CT examinations were performed using 16-detector CT scanners (Philips Brilliance 16, Philips Medical Systems, or Toshiba Aquilion 16, Toshiba Medical Systems); 1.5 mm × 16 collimations were used, and images with a slice thickness of 2 mm and a gap of 1 mm were reconstructed using a standard reconstruction algorithm. The tube voltage

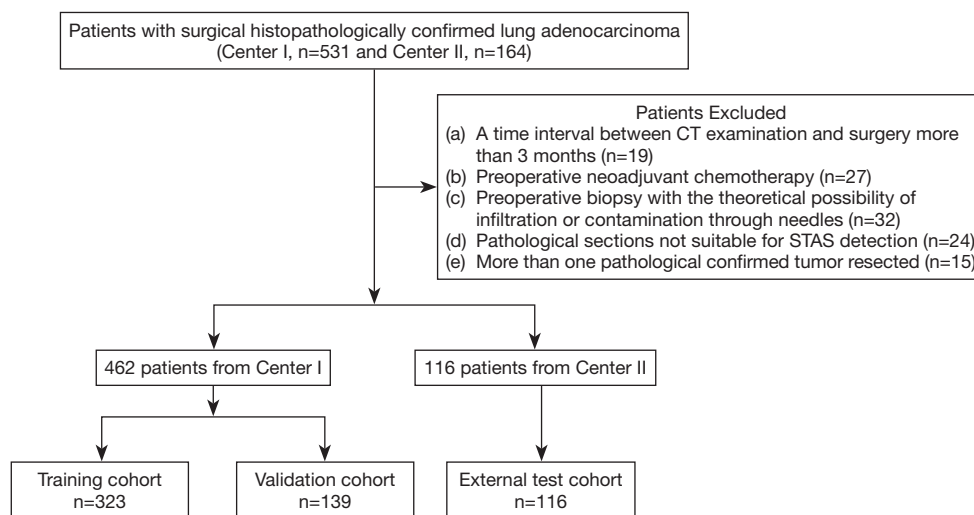


Figure 1 Flowchart of the study. STAS, spread through air space.

was 120 kV, and the tube current was automatically adjusted. Two experienced radiologists analyzed the CT images independently on the picture archiving and communication system (PACS) with a lung window (1,500 HU window width and -600 HU window level) and mediastinum window (250 HU window width and 40 HU window level). The interpretations included 3 measurements (maximum diameter, maximum area, and area of the solid components of the tumor) and 20 semantic characteristics (nodule type, pleural thickening, pleural retraction, mediastinal lymph node enlargement, hilar lymphatic enlargement, vascular cluster sign, lobulation, specula, air bronchogram, satellite lesions, vacuolar sign, void sign, pleural effusion, distribution within the lobe, nodules location, low central attenuation, other pulmonary nodules, emphysema, pulmonary fibrosis, and calcification). Measurements were performed on the transverse section that displayed the largest nodule using the lung window setting. The nodule area was measured by a manual drawing of the region of interest. For part solid nodules, the area of the solid component was measured, and the ratio of the solid components was calculated as the area of the solid component divided by the nodule area. Any disagreement was resolved through consensus, and the measurements were averaged.

Statistical analysis

The statistical analysis was performed in SPSS and R version 3.5.1 (R Foundation for Statistical Computing). A P value of less than 0.05 indicated statistical significance. Non-normal

distributed continuous variables are presented as medians and interquartiles. Categorical variables are presented as frequencies and percentages. At first, patients from center I were randomly divided into a training cohort (n=323) and a validation cohort (n=139) with a ratio of 0.7:0.3. Univariate analysis was used to select the CT characteristics associated with STAS in the patient cohort from center I. Measurements were compared with the Student's *t*-test or Mann-Whitney U test. Proportions were compared using either the χ^2 test or Fisher's exact test. Then, multivariable logistic regression was used to develop the CT characteristics on those variables with statistical significance. The model was validated in the validation cohort and then tested in the external test cohort (patients from center II). During the development of the model, we used the 'bestglm' package in R for automated parameter tuning with ten-fold cross-validation to find the optimal model. The diagnostic performance of the model was measured by area under the curve (AUC) of receiver operating characteristic (ROC). The AUC of the model for the validation cohort and the external test cohort was compared with the DeLong test using the 'pROC' package in R (13).

Results

Of the 462 tumors from the center I, 90 (19.5%) were found to be STAS positive. In the patient cohort from center II, STAS-positive nodules were identified in 28 patients (24.1%). The difference in the prevalence of STAS-positive nodules between the two centers was not statistically

Table 1 Clinical characteristics of the patient cohort from center I and II

Characteristics	Center I: STAS status			Center II: STAS status		
	Positive (n=90)	Negative (n=372)	P value	Positive (n=28)	Negative (n=88)	P value
Age (years)	55.1±13.2	58.8±12.4	0.013	53.4±13.7	61.2±10.9	0.018
Gender, n (%)			0.165			0.822
Male	46 (51.1)	160 (43.0)		13 (46.4)	43 (48.9)	
Female	44 (48.9)	212 (57.0)		15 (53.6)	45 (51.1)	
Smoker status, n (%)			0.075			0.936
Smoking	30 (33.3)	90 (24.2)		9 (32.1)	29 (33.0)	
Nonsmoking	60 (66.7)	282 (75.8)		19 (67.9)	5 (67.0)	

STAS, spread through air spaces.

significant ($\chi^2=1.238$, $P=0.266$). In center I cohort patients with STAS-positive tumors, it was found that the individuals were significantly younger than those with STAS-negative tumors. However, the differences in gender and smoking status between STAS-positive and STAS-negative tumors were not statistically significant (*Table 1*). Univariate analysis of the CT characteristics of STAS-positive and -negative nodules from centers I and II are summarized in *Table 2*. Eleven CT characteristics were associated with statistical significance. These factors included the maximum diameter of the tumor, the maximum area of the tumor, the area and ratio of the solid component, nodule type, pleural thickening, pleural retraction, mediastinal lymph node enlargement, vascular cluster sign, lobulation, and specula.

Regarding measurements, STAS-positive tumors tended to be larger in maximum diameter [18.23 mm (11.59, 28.12) *vs.* 24.53 mm (18.99, 37.51)], maximum area [172.04 (75.68, 426.30) *vs.* 385.02 (222.47, 719.88) mm²], areas of solid components [68.50 (0.00, 297.72) *vs.* 336.88 (193.56, 674.67) mm²], and had a higher ratio of solid components (85.17%±30.04% *vs.* 52.64%±43.15%).

As for the semantic characteristics, STAS-positive nodules tended to be solid or part-solid [ground glass nodules (GGNs), part solid, and solid were 3.3%, 18.9%, and 77.8% respectively, in STAS-positive nodules, and were 26.6%, 33.3%, and 40.1% respectively, in STAS-negative nodules]. STAS was also significantly associated with pleural thickening, pleural retraction, mediastinal lymph node enlargement, vascular cluster sign, lobulation, and specula.

CT images and histopathological photos of STAS-positive and STAS-negative nodules are shown in *Figures 2* and *3*.

The optimal logistic regression model included age, maximum diameter, and ratio of solid component with OR value of 0.967 (95% CI: 0.944–0.988), 1.027 (95% CI: 1.008–1.046), and 5.14 (95% CI: 2.180–13.321), respectively. This model obtained the AUC of 0.801 (95% CI: 0.709–0.892) and 0.692 (95% CI: 0.518–0.866) in the validation cohort and the external test cohort, respectively (*Figure 4*). The difference was not statistically significant ($P=0.280$).

Discussion

Our study showed that several CT characteristics of lung adenocarcinoma were associated with STAS, which implies that the radiologists' visual interpretation might have the potential to assess the STAS status preoperatively. Using automated parameter tuning, an optimal model which includes age, maximum diameter, and the ratio of the solid component can achieve high diagnostic performance for preoperative prediction of STAS in both the internal validation cohort and the external test cohort with AUC of 0.801 (95% CI: 0.709–0.892) and 0.692 (95% CI: 0.518–0.866) in the validation cohort and the external test cohort, respectively.

Widespread usage of low dose CT screening and micro intervention operation leads to the early detection of lung cancer and limited resection which preserves more lung parenchyma and improves patients' prognosis in turn (14). With the aid of artificial intelligence, CT scanning can be used to predict the invasive patterns of lung cancer (15). The newly defined invasion manner of lung adenocarcinoma, STAS, was shown to be a significant prognosticator for locoregional recurrence when patients with STAS-positive nodules received limited resection. However, if these

Table 2 Association between STAS and CT characteristics

Characteristics	Center I				Center II			
	Nodules (n=462)	STAS status		P value	Nodules (n=116)	STAS status		P value
		Positive (n=90)	Negative (n=372)			Positive (n=28)	Negative (n=88)	
Maximum diameter (mm)	20.00 (12.70, 30.20)	18.23 (11.59, 28.12)	24.53 (18.99, 37.51)	<0.001	23.40 (16.60, 37.63)	37.78 (20.78, 49.23)	21.98 (15.13, 33.46)	0.002
Maximum area (mm ²)	223.69 (89.67, 474.77)	172.04 (75.68, 426.30)	385.02 (222.47, 719.88)	<0.001	302.295 (150.30, 781.53)	727.80 (284.07, 1,239.99)	251.83 (130.20, 626.61)	0.001
Area of the solid component (mm ²)	114.08 (10.25, 384.46)	68.50 (0.00, 297.72)	336.88 (193.56, 674.67)	<0.001	254.91 (52.97, 705.17)	585.88 (220.75, 1,160.43)	181.40 (49.70, 523.56)	0.001
Ratio of solid component (%)	58.97±42.88	85.17±30.04	52.64±43.15	<0.001	69.89±37.20	79.67±32.25	60.77±38.28	0.014
Nodule type, n (%)				<0.001				0.096
GGN	102 (22.1)	3 (3.3)	99 (26.6)		11 (9.5)	0 (0.0)	11 (12.5)	
Part-solid	141 (30.5)	17 (18.9)	124 (33.3)		41 (35.3)	9 (32.1)	32 (36.4)	
Solid	219 (47.4)	70 (77.8)	149 (40.1)		64 (55.2)	19 (67.9)	45 (51.1)	
Pleural thickening, n (%)				0.002				0.186
Present	287 (62.1)	69 (76.7)	218 (58.6)		84 (72.4)	23 (82.1)	61 (69.3)	
Absent	175 (37.9)	21 (23.3)	154 (41.4)		32 (27.6)	5 (17.9)	27 (30.7)	
Pleural retraction, n (%)				0.001				0.142
Present	275 (59.5)	67 (74.4)	208 (55.9)		78 (67.2)	22 (78.6)	56 (63.6)	
Absent	187 (40.5)	23 (25.6)	164 (44.1)		38 (32.8)	6 (21.4)	32 (36.4)	
Mediastinal lymph node enlargement, n (%)				0.002				0.350
Present	54 (11.7)	19 (21.1)	35 (9.4)		22 (19.0)	7 (25.0)	15 (17.0)	
Absent	408 (88.3)	71 (78.9)	337 (90.6)		94 (81.0)	21 (75.0)	73 (83.0)	
Hilar lymphatic enlargement, n (%)				0.058				0.277
Absent	51 (11.0)	15 (16.7)	36 (9.7)		21 (18.1)	7 (25.0)	14 (15.9)	
Present	411 (89.0)	75 (83.3)	336 (90.3)		95 (81.9)	21 (75.0)	74 (84.1)	
Vascular cluster sign, n (%)				<0.001				0.162
Present	161 (34.8)	46 (51.1)	115 (30.9)		45 (38.8)	14 (50.0)	31 (35.2)	
Absent	301 (65.2)	44 (48.9)	257 (69.1)		71 (61.2)	14 (50.0)	57 (64.8)	
Lobulation, n (%)				<0.001				0.274
Present	205 (44.4)	55 (61.1)	150 (40.3)		60 (51.7)	17 (60.7)	43 (48.9)	
Absent	257 (55.6)	35 (38.9)	222 (59.7)		56 (48.3)	11 (39.3)	45 (51.1)	
Specula, n (%)				<0.001				0.590
Present	234 (50.3)	62 (68.9)	172 (46.2)		57 (49.1)	15 (53.6)	42 (47.7)	
Absent	228 (49.4)	28 (31.1)	200 (53.8)		59 (50.9)	13 (46.4)	46 (52.3)	
Air bronchogram, n (%)				0.912				0.446
Present	93 (20.1)	23 (25.6)	70 (18.8)		27 (23.3)	8 (28.6)	19 (21.6)	
Absent	369 (79.9)	67 (74.4)	302 (81.2)		89 (76.7)	20 (71.4)	69 (78.4)	
Satellite lesions, n (%)				0.988				0.501
Present	46 (10.0)	9 (10.0)	37 (9.9)		20 (17.2)	6 (21.4)	14 (15.9)	
Absent	416 (90.0)	81 (90.0)	335 (90.1)		96 (82.8)	22 (78.6)	74 (84.1)	

Table 2 (continued)

Table 2 (continued)

Characteristics	Center I			P value	Center II			P value
	Nodules (n=462)	STAS status			Nodules (n=116)	STAS status		
		Positive (n=90)	Negative (n=372)			Positive (n=28)	Negative (n=88)	
Vacuolar sign, n (%)				0.437				0.706
Present	99 (21.4)	22 (24.4)	77 (20.7)		26 (22.4)	7 (25.0)	19 (21.6)	
Absent	363 (78.6)	68 (75.6)	295 (79.3)		90 (77.6)	21 (75.0)	69 (78.4)	
Void sign, n (%)				0.513				0.806
Present	29 (6.3)	7 (7.8)	22 (5.9)		15 (12.9)	4 (14.3)	11 (12.5)	
Absent	433 (93.7)	83 (92.2)	350 (94.1)		101 (87.1)	24 (85.7)	77 (87.5)	
Pleural effusion, n (%)				0.383				0.011
Present	14 (3.0)	4 (4.4)	10 (2.7)		2 (1.7)	2 (7.1)	0 (0.0)	
Absent	448 (97.0)	86 (95.6)	362 (97.3)		114 (98.3)	26 (92.9)	88 (100.0)	
Distribution within the lobe, n (%)				0.334				0.321
Center	91 (19.7)	69 (76.7)	302 (81.2)		98 (84.5)	22 (78.6)	76 (86.4)	
Peripheral	371 (80.3)	21 (23.3)	70 (18.8)		18 (15.5)	6 (21.4)	12 (13.6)	
Nodule location, n (%)				0.415				0.239
LUL	116 (25.1)	27 (30.0)	89 (23.9)		29 (25.0)	11 (39.3)	18 (20.5)	
LLL	65 (14.1)	16 (17.8)	49 (13.2)		19 (16.4)	3 (10.7)	16 (18.2)	
RUL	164 (35.5)	26 (28.9)	138 (37.1)		43 (37.1)	7 (25.0)	36 (40.9)	
RML	33 (7.1)	5 (5.6)	28 (7.5)		6 (5.2)	2 (7.1)	4 (4.5)	
RLL	84 (18.2)	16 (17.8)	68 (18.3)		19 (16.4)	5 (17.9)	14 (15.9)	
Central low attenuation, n (%)				0.201				0.321
Present	32 (6.9)	9 (10.0)	23 (6.2)		18 (15.5)	6 (21.4)	12 (13.6)	
Absent	430 (93.1)	81 (90.0)	349 (93.8)		98 (84.5)	22 (78.6)	76 (86.4)	
Other pulmonary nodules, n (%)				0.819				0.963
Present	236 (51.1)	45 (50.0)	191 (51.3)		70 (60.3)	17 (60.7)	53 (60.2)	
Absent	226 (48.9)	45 (50.0)	181 (48.7)		46 (39.7)	11 (39.3)	35 (39.8)	
Emphysema, n (%)				0.903				0.985
Present	79 (17.1)	15 (16.7)	64 (17.2)		25 (21.6)	6 (21.4)	19 (21.6)	
Absent	383 (82.9)	75 (83.3)	308 (82.8)		91 (78.4)	22 (78.6)	69 (78.4)	
Pulmonary fibrosis, n (%)				0.406				0.921
Present	41 (8.9)	10 (11.1)	31 (8.3)		20 (17.2)	5 (17.9)	15 (17.0)	
Absent	421 (91.1)	80 (88.9)	341 (91.7)		96 (82.8)	23 (82.1)	73 (83.0)	
Calcification, n (%)				0.912				0.967
Present	11 (2.4)	2 (2.2)	9 (2.4)		4 (3.4)	1 (3.6)	3 (3.4)	
Absent	451 (97.6)	88 (97.8)	363 (97.6)		112 (96.6)	27 (96.4)	85 (96.6)	

Distribution within the lobe: center, the nodule entirely located within the inner two-thirds of the lung; peripheral, a portion of the nodule residing in the outer third of the lung. STAS, spread through air space; LUL, left upper lobe; LLL, left lower lobe; RUL, right upper lobe; RML, right middle lobe; RLL, right lower lobe.

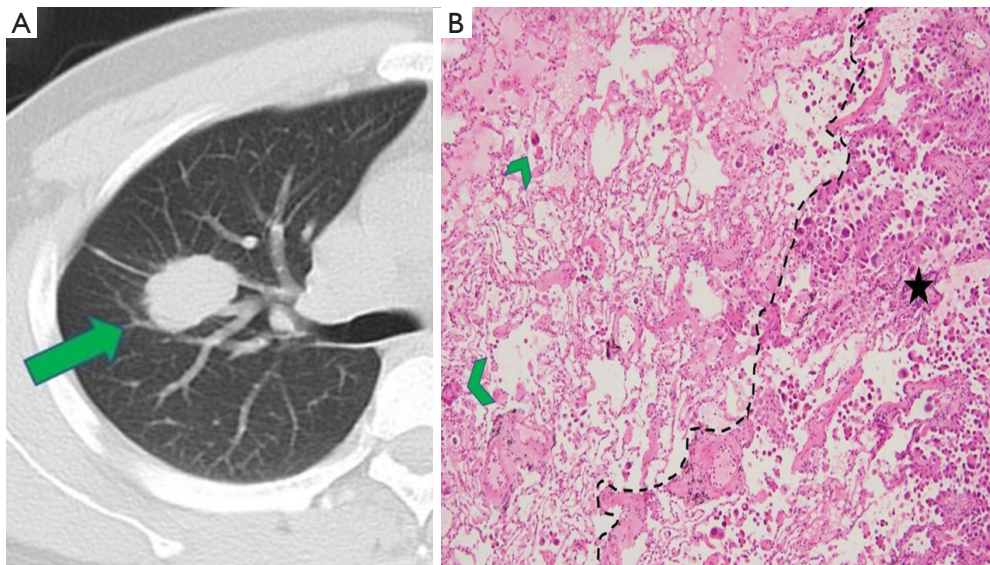


Figure 2 A 39-year-old man with STAS positive lung adenocarcinoma. (A) Axial CT image (width, 1,500 HU; level, -600 HU) showing a solid nodule of the right upper lobe (arrow). (B) Photomicrograph (hematoxylin-eosin stain, magnification $\times 200$) showing detached micropapillary clusters of tumor cells (arrowheads) in alveolus beyond the edge (dashed line) of the primary tumor (star).

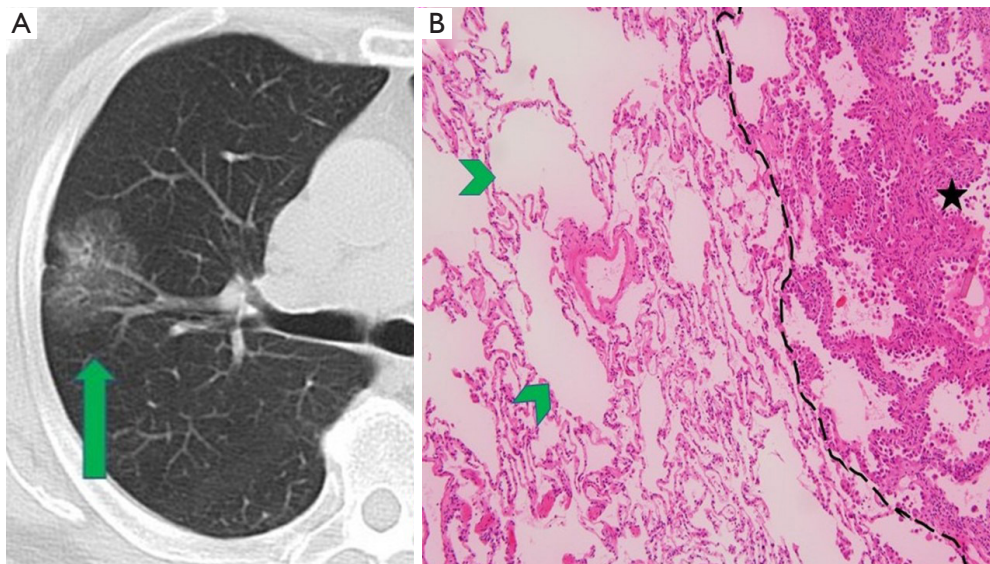


Figure 3 A 80-year-old man with STAS-negative lung adenocarcinoma. (A) Axial CT image (width, 1,500 HU; level, -600 HU) showing a ground-glass nodule of the right upper lobe (long arrow). (B) Photomicrograph (hematoxylin-eosin stain, magnification $\times 200$) showing clean alveolar spaces (arrowheads) adjacent to the boundary (dashed line) of the tumor (star).

patients underwent lobectomy, the association between STAS and tumor recurrence and overall patient survival was not observed (5). Therefore, preoperative knowledge of STAS status is essential for surgeons to choose the optimal operative type. STAS is a histopathological finding which

can only be discerned after operation. Before the operation, the surgeons cannot obtain information about the STAS status of lung nodules.

Several studies, including our previous research, attempted to investigate the association between CT

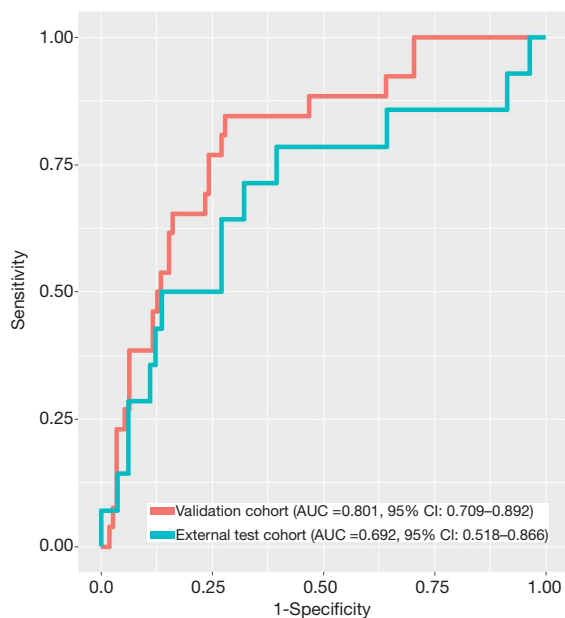


Figure 4 ROC curve of a CT-based logistic regression machine learning model for predicting STAS in lung adenocarcinoma in the validation cohort and the external test cohort. ROC, receiver operating characteristic; STAS, spread through air space.

features and STAS status or to predict STAS status using CT-based radiomics machine learning model (9-12). de Margerie-Mellon and colleagues noticed that size and a high proportion of solid components were associated with STAS (9). Using CT imaging features to develop a multivariable prediction model, Kim *et al.* found that the percentage of solid components was an independent predictor of STAS, which could achieve a sensitivity of 89.2% and a specificity of 60.3% when the cutoff value was set at 90% (11). Due to the low prevalence of STAS in early stage lung adenocarcinomas, both studies selected STAS-negative nodules with matched age, gender, and smoking status for analysis. Therefore, selection bias was introduced. In the present study, age was found to be an independent risk factor, and young patients tended to have STAS positive tumors. In the study performed by Toyokawa *et al.*, presence of notch and the absence of GGO were demonstrated to be independent risk factors for the STAS phenomenon. Although Toyokawa *et al.* proved that the ratio of the solid components at four levels was different between STAS-positive and -negative tumors, they did not quantify the solid components in the multivariate analysis. Therefore, they might have exaggerated the contribution of these nodule types. In this study, the univariate analysis

showed that both the nodule type and the ratio of the solid components were associated with STAS status, while multivariate analysis revealed that only the ratio of solid components was the risk factor. Nodule type and the ratio of solid components were related to the variables, and the ratio of solid components could quantify the sub-solid nodules. Our previous study showed that the median, maximum, age, maximum 3D diameter, and size-zone non-uniformity normalized scale were listed as the top five critical radiomics features with a random forest machine learning model, which obtained an AUC of 0.75 (12). Our study is the first to develop and validate a model to predict STAS on CT characteristics. Three variables, which consisted of age, maximum diameter, and the ratio of solid components, were selected to develop an optimal model using the leveraging automated parameter tuning R package, ‘bestglm’. The two CT characteristics included in the model were concordant with what had been reported by de Margerie-Mellon and Kim. Nodules with larger size and more solid components tended to be STAS positive. Adding age, our model performed as well as the radiomics model.

Furthermore, we added an external test cohort, which was unseen by the model. Although the only moderate performance was obtained in the test cohort, it reflected the model’s ability to predict STAS in the real world. Therefore, our study addressed not only the association between CT characteristics and STAS status of lung adenocarcinomas, but the possibility to harness perceptible CT characteristics to predict STAS, which allows it to be a routinely available modality in the real world and manages to eliminate the need for time-consuming segmentation.

There are some limitations to this study. First, STAS is a histopathological finding, and so we could only enroll those patients who had operation to resect tumors. The nodules tended to be large, and had more solid components or a greater likelihood of being malignant at visual interpretation. Therefore, selection bias might have been introduced. Second, only 19.5% nodules in the center I cohort and 24.1% nodules in the center II cohort were STAS positive. This imbalanced data could have decreased the accuracy of the model. Augmentation or down-sampling was conducted to improve the model’s performance and prevent overfitting. Third, the lack of follow-up prevented us from evaluating the effects of STAS on patients’ prognosis.

In conclusion, at univariate analysis, several CT characteristics were associated with the STAS status of lung adenocarcinomas. Age, maximum diameter, and the ratio of solid components were selected by the ‘bestglm’ package in

R to develop an optimal model for predicting STAS status. The model showed good diagnostic performance in both the internal validation cohort and the external test cohort, which demonstrated that routine visual interpretation of CT images could be useful for preoperative assessment of STAS status for lung adenocarcinomas.

Acknowledgments

Funding: The Shenzhen Science and Technology Project supported this study (No. GJHZ20180928172002087).

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <http://dx.doi.org/10.21037/qims-20-724>). JG serves as an unpaid editorial board member of *Quantitative Imaging in Medicine and Surgery*. The other authors have no conflicts of interest to declare.

Ethical Statement: This retrospective study was given approval by the institutional review boards of both centers, and the requirement for informed consent was waived.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JHM, Beasley MB, Chirieac LR, Dacic S, Duhig E, Flieder DB, Geisinger K, Hirsch FR, Ishikawa Y, Kerr KM, Noguchi M, Pelosi G, Powell CA, Tsao MS, Wistuba I; WHO Panel. The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification. *J Thorac Oncol* 2015;10:1243-60.
2. Amin MB, Tamboli P, Merchant SH, Ordóñez NG, Ro J, Ayala AG, Ro JY. Micropapillary component in lung adenocarcinoma: a distinctive histologic feature with possible prognostic significance. *Am J Surg Pathol* 2002;26:358-64.
3. Blaauwgeers H, Flieder D, Warth A, Harms A, Monkhorst K, Witte B, Thunnissen E. A Prospective Study of Loose Tissue Fragments in Non-Small Cell Lung Cancer Resection Specimens: An Alternative View to "Spread Through Air Spaces". *Am J Surg Pathol* 2017;41:1226-30.
4. Onozato ML, Kovach AE, Yeap BY, Morales-Oyarvide V, Klepeis VE, Tammireddy S, Heist RS, Mark EJ, Dias-Santagata D, Iafrate AJ, Yagi Y, Mino-Kenudson M. Tumor islands in resected early-stage lung adenocarcinomas are associated with unique clinicopathologic and molecular characteristics and worse prognosis. *Am J Surg Pathol* 2013;37:287-94.
5. Kadota K, Nitadori JI, Sima CS, Ujii H, Rizk NP, Jones DR, Adusumilli PS, Travis WD. Tumor Spread through Air Spaces is an Important Pattern of Invasion and Impacts the Frequency and Location of Recurrences after Limited Resection for Small Stage I Lung Adenocarcinomas. *J Thorac Oncol* 2015;10:806-14.
6. Dai C, Xie H, Su H, She Y, Zhu E, Fan Z, Zhou F, Ren Y, Xie D, Zheng H, Kadeer X, Chen D, Zhang L, Jiang G, Wu C, Chen C. Tumor Spread through Air Spaces Affects the Recurrence and Overall Survival in Patients with Lung Adenocarcinoma >2 to 3 cm. *J Thorac Oncol* 2017;12:1052-60.
7. Berfield KS, Wood DE. Sublobar resection for stage IA non-small cell lung cancer. *J Thorac Dis* 2017;9:S208-S210.
8. Warth A, Muley T, Kossakowski CA, Goepfert B, Schirmacher P, Dienemann H, Weichert W. Prognostic Impact of Intra-alveolar Tumor Spread in Pulmonary Adenocarcinoma. *Am J Surg Pathol* 2015;39:793-801.
9. de Margerie-Mellon C, Onken A, Heidinger BH, VanderLaan PA, Bankier AA. CT Manifestations of Tumor Spread Through Airspaces in Pulmonary Adenocarcinomas Presenting as Subsolid Nodules. *J Thorac Imaging* 2018;33:402-8.
10. Toyokawa G, Yamada Y, Tagawa T, Kamitani T, Yamasaki Y, Shimokawa M, Oda Y, Maehara Y. Computed tomography features of resected lung adenocarcinomas with spread through air spaces. *J Thorac Cardiovasc Surg* 2018;156:1670-1676.e4.
11. Kim SK, Kim TJ, Chung MJ, Kim TS, Lee KS, Zo JI, Shim YM. Lung Adenocarcinoma: CT Features Associated with Spread through Air Spaces. *Radiology* 2018;289:831-40.
12. Jiang C, Luo Y, Yuan J, You S, Chen Z, Wu M, Wang G, Gong J. CT-based radiomics and machine learning to

- predict spread through air space in lung adenocarcinoma. *Eur Radiol* 2020;30:4050-7.
13. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837-45.
 14. Mao L, Chen H, Liang M, Li K, Gao J, Qin P, Ding X, Li X, Liu X. Quantitative radiomic model for predicting malignancy of small solid pulmonary nodules detected by low-dose CT screening. *Quant Imaging Med Surg* 2019;9:263-72.
 15. Wang S, Wang R, Zhang S, Li R, Fu Y, Sun X, Li Y, Sun X, Jiang X, Guo X, Zhou X, Chang J, Peng W. 3D convolutional neural network for differentiating pre-invasive lesions from invasive adenocarcinomas appearing as ground-glass nodules with diameters ≤ 3 cm using HRCT. *Quant Imaging Med Surg* 2018;8:491-9.

Cite this article as: Li C, Jiang C, Gong J, Wu X, Luo Y, Sun G. A CT-based logistic regression model to predict spread through air space in lung adenocarcinoma. *Quant Imaging Med Surg* 2020;10(10):1984-1993. doi: 10.21037/qims-20-724