# Comparison between semantic features and lung-RADS in predicting malignancy of screening lung nodule

**Qian Li**[1,2], **Yoganand Balagurunathan**[2], **Ying Liu**[1], **Jin Qi**[1,2], **Matthew B. Schabath**[3], **Zhaoxiang Ye**[1,*], and **Robert Gillies**[2,*]

[1]Department of Radiology, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center of Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin, China.

[2]Department of Cancer Imaging and Metabolism, H. Lee Moffitt Cancer Center, Tampa, FL, USA.

[3]Department of Cancer Epidemiology, H. Lee Moffitt Cancer Center, Tampa, FL, USA.

## Abstract

**Rationale**—Lung-RADS is proposed for the Low-dose computed tomography (LDCT) interpretation in lung cancer screening, but its performance needs to be further evaluated.

**Objectives**—To compare the value of radiological semantic features and lung-RADS in predicting nodule malignancy risk at different screening rounds, and to investigate whether the predictive power of lung-RADS could be improved by incorporating semantic features.

**Methods**—A training cohort of 199 patients (139 benign and 60 cancerous nodules diagnosed at the third screening round), and a testing cohort of 80 patients (40 benign and 40 malignant nodules) were obtained from the National Lung Screening Trial dataset. A multivariate linear predictor model was built based on the 24 systematically scored semantic features, and the performances were compared to lung-RADS (scale 3 or above called positive).

**Measurements and Main Results**—Among the semantic features, contour and border definition were the top individual predictors. The average area under the receiver-operating characteristic curve (AUC) of border definition at baseline (T0) was 0.724. The average AUC of contour at first (T1) and second follow-up (T2) were 0.843 and 0.878, respectively. Other significant features included size, location, vessel attachment, solidity, focal emphysema and focal

\*Corresponding author. Robert J. Gillies, PhD, H. Lee Moffitt Cancer Center and Research Institute, 12902 Magnolia Drive, Tampa, FL 33612. Tel: 1-813-745-8355, Fax: 813-745-7265; robert.gillies@moffitt.org, Zhaoxiang Ye, MD, Tianjin Medical University Cancer Institute and Hospital, Huan-Hu-Xi Road, Ti-Yuan-Bei, He Xi District, Tianjin 300060, China. Fax: +86-22-23537796; yezhaoxiang@163.com.

fibrosis. In comparison, the average AUC of lung-RADS at T0, T1 and T2 were 0.600, 0.760 and 0.867, respectively, and could be improved to 0.743, 0.887 and 0.968 by adding semantic features.

**Conclusion—**The semantic features performed similar to lung-RADS at follow-ups, outperformed lung-RADS at baseline, and could improve the performance of lung-RADS for all screening rounds.

### Keywords

semantic features; lung-RADS; NLST; predictive

## 1. Introduction

Lung cancer is the leading cause of cancer death both in the US and the rest of the world (1–3). Until recently, no screening method has been shown to help decreasing lung cancer mortality rates. The National Lung Screening Trial (NLST) found a 20% reduction in lung cancer mortality for participants screened with low-dose computed tomography (LDCT) versus standard chest radiography (4). However, high-false positive rates and overdiagnosis are limitations of screening with LDCT. Across the three rounds of screening in NLST, 96.4% of the positive detections were not cancerous (4), as evaluated with further imaging or invasive procedures. The consequence of the false detection includes increased use of medical resources, additional radiation exposure, complications arising from invasive procedures and patient anxiety. Therefore, it is critical to develop reliable image markers that can predict nodule malignancy.

In the NLST, a positive screen criterion was based solely on nodule size, that is, a non-calcified nodule of 4mm or greater in any diameter. Subsequent studies have proposed that raising the threshold to 5 to 9mm would substantially reduce false positive rates at the expense of a few missed or delayed lung cancer cases (5–7). There are a few diagnostic models (8–12) that propose to add other image features, such as nodule solidity, location and emphysema, and these showed improved malignancy detection. Nevertheless, most of these studies focused on differentiating between benign and malignant nodules, regardless of predicting malignant progression.

Recently, the American College of Radiology (ACR) proposed lung imaging reporting and data system (lung-RADS) for LDCT screening interpretation (13), which adopts different criteria for baseline and subsequent scans. This system combines nodule size and solidity to provide risk assessment, and uses different size thresholds based on solidity, that is, an average diameter of 20 mm for ground-glass nodules, 6 mm for solid nodules, and 6mm total diameter for part-solid nodules. Though previous studies using lung-RADS (14, 15) showed a reduction of false-positive rate with the assessment of both baseline and subsequent scans, lack of nodule tracking throughout the subsequent follow-up scans (14) may lead to inaccurate evaluation of the performance of lung-RADS. It could be that the nodule assessed in the baseline scan was not consistent with the one that analyzed during follow-up scans.

In this analysis, both baseline and follow-up LDCT scans were extracted from NLST dataset and all these images were reviewed to ensure that the nodules analyzed were matched across all three rounds of screening. A systematic radiological scoring sheet (semantics), which included nodule size, shape, margin, density, internal features, external features and associate findings, were developed with the goal of finding reliable semantic descriptors that will not only help in predicting malignancy, but also predict whether the nodule would develop cancer in the future. The performances of top semantic features were then compared with lung-RADS, and investigate whether they could improve the prediction accuracy of lung-RADS.

## 2. Patients and methods

### 2.1 NLST Study Population

The NLST dataset was obtained after the Data transfer agreement (DTA) between the National Cancer Institute (NCI) and Moffitt Cancer Center (MCC). The Institutional Review Board (IRB) at MCC allows retrospective data analysis of publically available data. The image and clinical data was accessed through a data portal supported by the data managers for the NLST (16). The detailed study design of the NLST have been described in a previous paper (4). A total of 53,454 subjects (between 55 and 74 years of age) with high-risk of lung cancer at the time of randomization were enrolled from 2002 through 2004 at 33 U.S medical centers. The participants were randomly assigned to the LDCT group and the radiography group and asked to undergo a baseline (T0) and two annual follow-up screenings (T1 and T2). Participants with lung cancer diagnosed would not be offered subsequent screening tests. All screening examinations were performed in accordance with a standard protocol.

In this study, we formed training and test sets that consisted of incidence lung cancer patients and nodule positive controls. In the training cohort, 92 patients had positive nodules not related with lung cancer diagnosis at baseline (T0) and first follow-up (T1), but confirmed to be cancer at the second follow-up (T2). In the test set there were 104 lung cancer patients that had a baseline (T0) scan and confirmed to be cancer at the first follow up (T1). Nodule positive controls had three consecutive scans (T0 to T2) with a benign nodule and were frequency matched 2:1 to the lung cancer cases on age (+/− 5 years), sex, race, smoking status, and pack-years smoked The schema of the study is shown in Fig. 1.

The cancerous nodule location was provided by the NLST, while the location of benign nodule was not available. Two radiologists (J.Q. and Y.L.) reviewed the images and reached an agreement on the nodules used for analysis and those nodules excluded from the analysis. One radiologist (J.Q.) reviewed all the images at three time points to make sure that the nodules involved in the analysis were consistent across the 3 screening rounds. About 77 cases in training cohort were excluded because of one of the following reasons: non-availability of T0, T1 or T2 time point images, the location of tumor was unknown, nodules cannot be identified, or multiple nodules that had malignant characteristics. In some cases, nodules were too small to be evaluated especially at baseline. At last, there were 199 cases (60 cancerous and 139 benign) qualified for the training cohort. Additionally, 80 patients were randomly chosen for the feature model validation, which had 40 patients with

confirmed cancer and 40 patients continuing to be benign at last follow up. The patients' IDs for both cohorts were listed in Supplementary Table 1.

## 2.2 LDCT images analysis

Image scans at all three time points (T0, T1 and T2) were analyzed. LDCT images were displayed using both mediastinal (width, 350 HU; level, 40 HU) and lung (width, 1500 HU; level, -600 HU) window settings. Totally, 25 radiological image traits were identified to characterize the pulmonary nodules. These semantics can be broadly classified into eight categories: (1) location; (2) size; (3) shape; (4) margin; (5) density; (6) internal features; (7) external features; and (8) associated findings (Supplementary Table 2). These features were systematically scored on a point scale (up to 5) by the radiologist (Q.L.). Lung-RADS scores were independently evaluated in each time point according to the ACR 2014 guidelines (13). To measure the reproducibility accuracy of scoring the semantics, we selected 40 patients (20 malignant, 20 benign in blinded fashion) from the NLST trial and provided the scoring sheet with approximate anatomical location of the nodules to another radiologists (Y.L.). Both of the two radiologists were blinded to the case-control status.

## 2.3 Statistical analysis

Agreement between the two readers (Y.L and Q.L) was measured by the (weighted) Kappa index (17, 18) for binary or ordinal variable and intra-class correlation of coefficient (ICC) (19) for continuous variable. The kappa value was interpreted as follows: < 0: less than chance agreement; 0.01 to 0.2: slight agreement; 0.21 to 0.4: fair agreement; 0.41 to 0.6: moderate agreement; 0.61 to 0.8: substantial agreement; > 0.8: almost perfect agreement (20).

Discriminatory analysis was conducted using a liner classifier to find the best pairs of predictive features that relate to cancer status. The error of classification was estimated using a 5-fold holdout cross validation method, randomized and repeated large number of times (over 200). We report the average statistics across the repeats. For each discriminant feature pair, AUC and 95% confidence interval (CI) was computed. An exhaustive search was used to find the best features in all possible feature pairs (up to fourth dimension, over 12,650 pairs). The top discriminating features was ordered based on the Youdon's J Index (sensitivity + specificity – 1) and the top discriminant pair was reported. The sensitivity, specificity, average AUC and 95% CI were also calculated for lung-RADS (scale 3 or above called positive). The discriminatory analysis was repeated independently at each of the time points. The performance of the top features was tested on the validation cohort.

## 3. Results

The demographic and clinical covariates of the training and test sets are provided in Table 1. Most of the lung cancer patients were in stage I (training: 76.6%, testing: 72.5%) when diagnosis were confirmed, and adenocarcinoma was the main histological subtype (training: 68.4%, testing: 70%).

### 3.1 Performance of lung-RADS in NLST

Overall, lung-RADS performed better when a prior scan was available (i.e. at subsequent screenings) compared to its performance at baseline. At the second follow-up scan (T2), when nodule malignancy was confirmed, the AUC of lung-RADS in discriminating benign and malignant nodules was 0.867, with the sensitivity of 0.782 to 0.856, specificity of 0.947 to 0.964 (table 2c). At T1 (table 2b), the ability of lung-RADS in predicting whether the nodule would subsequently present as a clinically-relevant cancer one year later was moderate, the AUC, sensitivity and specificity were 0.760, 0.537 – 0.621 and 0.949 – 0.970, respectively. At baseline screening (T0, table 2a), lung-RADS showed an AUC of 0.600, and the sensitivity is relatively lower, in the range of [0.399 – 0.466], specificity was [0.750 – 0.794].

### 3.2 Performance of semantic CT features in NLST

Two features, distribution and calcification, were excluded because most of the nodules were peripheral and non-calcified in the study. Among the remaining features, ten showed almost perfect agreement (Kappa value > 0.8), including location, vessel attachment, solidity, air bronchogram, fissure attachment, pleural attachment, pleural retraction, bubble-like lucency, thickened adjacent bronchovascular bundles and nodules in primary tumor lobe, while contour, border definition, concavity, lymphadenopathy, spiculation, nodules in non-tumor lobes and focal fibrosis were in substantial agreement, and focal emphysema, vascular convergence and lobulation showed moderate agreement. The ICCs for long and short axial diameter were 0.940 (95% CI: 0.890 – 0.968) and 0.960 (95% CI: 0.834 – 0.985), respectively. Detailed information can be seen in Table 3.

As shown in Table 2, the performance of semantic features was evaluated both individually and by combining with others, such as two, three or four feature combinations (also called feature dimensions). The comparison among three screening rounds (T0, T1 and T2) showed that the sensitivity of semantic features increased with screening interval. The sensitivity at T0 is about 0.5 – 0.6, at T1 it is about 0.6 – 0.7, and at T2 it is around 0.8. The specificity had small changes across different time points, which was in the range of 0.8 – 0.9. The comparison among different feature dimensions (1D, 2D, 3D and 4D) in each screening rounds showed that the feature dimension plays a role in improving predictor performance (AUC). For example, at T2 (Table 2c), location, contour, short axial diameter and focal emphysema showed the highest AUC of 0.947.

In all three screening rounds, the size of benign nodules was significantly smaller than the malignant nodules ($P < 0.05$) in the training cohort (Supplementary Table 3). The nodule growth rates between two sequential scans were also statistically different between cancer and normal groups, which are in line with the NLST selection criteria.

It was found that contour and border definition as individual semantic feature were predictive of malignancy in all three screening rounds (Figure 2), with border definition performed better at T0 (AUC: 0.719), and contour slightly better at T1 (AUC: 0.823) and T2 (AUC: 0.876). Vessel attachment was another prognostic feature in all three screening rounds, but it needed to be combined with other two features to generate sufficient power.

For example, when combined with contour and focal emphysema, it was prognostic both at T1 and T2; while combined with border definition and solidity, it was prognostic at T0. Similarly, location was a significant feature at T1 and T2, but has to be combined with contour. Besides these, at T0 (Table 2a), focal fibrosis were also predictive features if combined with border definition, and the combination, short axial diameter, contour, border definition and solidity, showed the highest AUC (0.741). At T1 (Table 2b), it was observed that location, contour, solidity and focal fibrosis ranked the top combinations (AUC: 0.897).

There is no obvious difference between semantic features and lung-RADS in diagnosing nodule malignancy at T2, but the AUC of semantic features at T1 and T0 were higher than that of lung-RADS. Then we tried to incorporate the semantic features into the predictive model to improve the performance of lung-RADS (Figure 3). At T2 and T1, the AUC increased about 10% for the top combined model compared to lung-RADS predictions (0.968 vs. 0.867, 0.887 vs 0.760). At baseline (T0), the AUC was as high as 0.743 (lung-RADS, 0.600) by incorporating border definition, and it was similar with the AUC of four feature semantic model (short axial diameter, contour, border definition and solidity, 0.741).

## 4. Discussion

Using data and images from the NLST, we performed a systemic analysis of radiological semantic features of incidence lung cancer cases and nodule positive controls and found that imaging features distinguished malignant from benign nodules and predicted subsequent tumor occurrence. The performance of the semantic features was comparable to that of lung-RADS for subsequent scans, and better for the baseline scan. By combining semantic features with lung-RADS, the ability of predicting nodule malignancy improved. It indicates that a quantitative semantics could be considered as surrogate to lung-RADS and in combination will certainly add information to identify malignant nodules.

### 4.1 Performance of Lung-RADS in NLST

The lung-RADS worked well in discriminating benign and malignant nodules when a previous scan was available, while less than expected in predicting nodule malignancy progression at baseline. The reason is that nodule size was too small to be defined as malignant at baseline. Meanwhile, it should be noted that lung-RADS relies on nodule growth for subsequent scans, and actual size is only used for baseline scans. To illustrate the difference, we tried to reclassify the lung-RADS category at T2 using the actual size, ignoring prior scan. The false positive for such a method increased about 20% (from 3.6% to 25.89%). it indicated that nodule growth works better than actual size in reducing false positive rate.

In comparison with Pinsky et al's study (14) that used lung-RADS in NLST data, the sensitivity and specificity after baseline were similar, but our baseline sensitivity is lower. The discrepancy may be attributed to the following: First, the sample size is different. They used all the subjects in NLST (LDCT cohort), while in our study we used subsets of LDCT cohort and divided them into train and validation/test cohorts. Also only patients with images of all three (two for test cohort) time points available could be included. Second, the ratios of malignant and benign nodules were significantly different. In our analysis, the

cancer and non-cancer cohort were matched at the ratio of 1:2, and the ratio in their study at baseline was nearly 1:90. Moreover, all the nodules analyzed were well matched across different screening rounds in our study. Though larger nodules may coexist, only the nodule which developed cancer at T2 was chosen, and it may result that some of the nodules at T0 were rather small (Figure 4). There were 24 patients (40%) smaller than 6mm in T0, which made the sensitivity lower.

## 4.2 Semantic features that can predict nodule malignancy

The Lung-RADS scoring system relies on the nodule size and solidity, while there are other sematic features that can be used as image biomarkers of malignancy. Among the semantic features, contour and border definition shows up as the top candidates that are related to malignancy. They can not only distinguish malignant from benign nodules, but also predict nodule malignancy progression using pre-diagnostic scans. This means that the malignant nodules tends to be more irregular and more poorly defined margins, similar observation are made in previous studies (21–25). It suggests that these semantic features play critical roles and should be evaluated in the screening setting, especially when the sensitivity of lung-RADS is modest (i.e at baseline scan).

Vessel invasion has shown to be a poor prognostic factor and is correlated with lung cancer recurrence (26–28). It has been demonstrated that during the early stages of tumor growth, angiogenesis is required to permit tumor expansion (29). Nodules with vessel attachment may be more easily to get vessel involved and develop progression.

Nodule location was another important prognostic indicator of malignancy (9) and it has been shown in multiple studies (24). Cancerous nodules are more often seen in the upper lobes, it is about 61.7% (n = 37) in our study. In comparison, 56 (40.3%) nodules in non-cancer cases have nodule located in the upper lobe.

COPD and interstitial lung disease are strong risk factors for lung cancer (30–32). It has been shown that even early signs of emphysema and fibrosis are associated with lung cancer (33) and regional severity of emphysema is an independent predictor of long-term survival (34). Henschke et al (35) found that the prevalence of lung cancer increased among those who had CT evidence of emphysema, no matter smokers or never smokers. In our study, malignant nodules tended to have severe regional emphysema and fibrosis. Based on prior study findings, visual assessment of emphysema and fibrosis on CT images could be used in lung cancer risk analysis.

## 4.3 Performance comparison between radiological features and lung-RADS

At T1 and T2 scan round, the performance of semantics defined by contour and border definition is comparable to that of lung-RADS. It is necessary to note that the lung-RADS score use prior scan information, whereas radiological semantic predictors use current scan to characterize a nodule. These semantic descriptors are common features that radiologist's use in daily practice, following semantic approaches saves time and effort.

At T0 scan round, the semantic features outperformed lung-RADS. Moreover, the predictive power of lung-RADS could be improved if semantic features were incorporated. It indicates

that the semantic features should be taken into account for cancer risk assessment especially when no previous scan was available.

### 4.4. Limitations

This study was based on retrospective analysis of patient data. Though we had assembled comparably large cohort for train and validation/test, this still remained small. Recently, quantitative descriptors of lung nodules (radiomics) has shown great potential for prognosis in lung cancer, which, along with semantics, could be used to improve malignancy prediction and improve false discovery rate in the future.

## Conclusion

In this study, we have shown radiological based discrimination approach is comparable to lung-RADS in classifying the presence of cancer, and holds value in using pre-diagnostic scans to predict subsequent occurrence of tumor. Evaluating semantic features along with lung-RADS could improve the screening cancer risk assessments at baseline. A quantitative radiological approach may act complementary assessment tool to the lung-RADS.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

## References

1. Torre LA, Siegel RL, Jemal A. Lung Cancer Statistics. Adv Exp Med Biol. 2016; 893:1–19. [PubMed: 26667336]

2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. CA: a cancer journal for clinicians. 2016; 66:7–30. [PubMed: 26742998]

3. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer. 2015; 136:E359–E386. [PubMed: 25220842]

4. National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med. 2011; 365:395–409. [PubMed: 21714641]

5. Henschke CI, Yip R, Yankelevitz DF, Smith JP. Definition of a positive test result in computed tomography screening for lung cancer: a cohort study. Ann Intern Med. 2013; 158:246–252. [PubMed: 23420233]

6. Gierada DS, Pinsky P, Nath H, Chiles C, Duan F, Aberle DR. Projected outcomes using different nodule sizes to define a positive CT lung cancer screening examination. J Natl Cancer Inst. 2014; 106:dju284. [PubMed: 25326638]

7. Yip R, Henschke CI, Yankelevitz DF, Smith JP. CT screening for lung cancer: alternative definitions of positive test result based on the National Lung Screening Trial and International Early Lung Cancer Action Program databases. Radiology. 2014; 273:591–596. [PubMed: 24955929]

8. Zhang M, Zhuo N, Guo Z, Zhang X, Liang W, Zhao S, He J. Establishment of a mathematic model for predicting malignancy in solitary pulmonary nodules. J Thorac Dis. 2015; 7:1833–1841. [PubMed: 26623107]

9. McWilliams A, Tammemagi MC, Mayo JR, Roberts H, Liu G, Soghrati K, Yasufuku K, Martel S, Laberge F, Gingras M. Probability of cancer in pulmonary nodules detected on first screening CT. New Engl J Med. 2013; 369:910–919. [PubMed: 24004118]

10. Li Y, Chen K-Z, Wang J. Development and validation of a clinical prediction model to estimate the probability of malignancy in solitary pulmonary nodules in Chinese people. Clin Lung Cancer. 2011; 12:313–319. [PubMed: 21889113]

11. Swensen SJ, Silverstein MD, Ilstrup DM, Schleck CD, Edell ES. The probability of malignancy in solitary pulmonary nodules: application to small radiologically indeterminate nodules. Arch Intern Med. 1997; 157:849–855. [PubMed: 9129544]

12. Gould MK, Ananth L, Barnett PG. A clinical model to estimate the pretest probability of lung cancer in patients with solitary pulmonary nodules. Chest. 2007; 131:383–388. [PubMed: 17296637]

13. American College of Radiology. Lung CT Screening Reporting and Data System (Lung-RADS™). 2014. Available from: http://www.acr.org/Quality-Safety/Resources/LungRADS

14. Pinsky DSG, Paul F., Black, William, Munden, Reginald, Nath, Hrudaya, Aberle, Denise, Kazerooni, Ella. Performance of Lung-RADS in the National Lung Screening Trial: a retrospective assessment. Ann Intern Med. 2015; 162:485–491. [PubMed: 25664444]

15. McKee BJ, Regis SM, McKee AB, Flacke S, Wald C. Performance of ACR Lung-RADS in a clinical CT lung screening program. J Am Coll Radiol. 2015; 12:273–276. [PubMed: 25176499]

16. National Cancer Institute: Cancer Data Access System. Available from: https://biometry.nci.nih.gov/cdas/studies/nlst/

17. Carletta J. Assessing agreement on classification tasks: the kappa statistic. Comput Linguist. 1996; 22:249–254.

18. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther. 2005; 85:257–268. [PubMed: 15733050]

19. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull. 1979; 86:420. [PubMed: 18839484]

20. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. Fam Med. 2005; 37:360–363. [PubMed: 15883903]

21. Xu DM, van der Zaag-Loonen HJ, Oudkerk M, Wang Y, Vliegenthart R, Scholten ET, Verschakelen J, Prokop M, de Koning HJ, van Klaveren RJ. Smooth or Attached Solid Indeterminate Nodules Detected at Baseline CT Screening in the NELSON Study: Cancer Risk during 1 Year of Follow-up. Radiology. 2009; 250:264–272. [PubMed: 18984780]

22. van't Westeinde SC, de Koning HJ, Xu D-M, Hoogsteden HC, van Klaveren RJ. How to deal with incidentally detected pulmonary nodules less than 10mm in size on CT in a healthy person. Lung Cancer. 2008; 60:151–159. [PubMed: 18359124]

23. Takashima S, Sone S, Li F, Maruyama Y, Hasegawa M, Kadoya M. Indeterminate solitary pulmonary nodules revealed at population-based CT screening of the lung: using first follow-up diagnostic CT to differentiate benign and malignant lesions. Am J Roentgenol. 2003; 180:1255–1263. [PubMed: 12704034]

24. Lindell RM, Hartman TE, Swensen SJ, Jett JR, Midthun DE, Tazelaar HD, Mandrekar JN. Five-year Lung Cancer Screening Experience: CT Appearance, Growth Rate, Location, and Histologic Features of 61 Lung Cancers. Radiology. 2007; 242:555–562. [PubMed: 17255425]

25. Ost D, Fein AM, Feinsilver SH. The solitary pulmonary nodule. New Engl J Med. 2003; 348:2535–2542. [PubMed: 12815140]

26. Tsuchiya T, Akamine S, Muraoka M, Kamohara R, Tsuji K, Urabe S, Honda S, Yamasaki N. Stage IA non-small cell lung cancer: vessel invasion is a poor prognostic factor and a new target of adjuvant chemotherapy. Lung Cancer. 2007; 56:341–348. [PubMed: 17350137]

27. Maeda R, Yoshida J, Ishii G, Hishida T, Nishimura M, Nagai K. Prognostic impact of intratumoral vascular invasion in non-small cell lung cancer patients. Thorax. 2010; 65:1092–1098. [PubMed: 20971984]

28. Ruffini E, Asioli S, Filosso PL, Buffoni L, Bruna MC, Mossetti C, Solidoro P, Oliaro A. Significance of the Presence of Microscopic Vascular Invasion After Complete Resection of Stage I–II pT1-T2N0 Non-small Cell Lung Cancer and Its Relation with T-Size Categories: Did the 2009 7th Edition of the TNM Staging System Miss Something? J Thorac Oncol. 2011; 6:319–326. [PubMed: 21164365]

29. Kessler R, Gasser B, Massard G, Roeslin N, Meyer P, Wihlm J-M, Morand G. Blood vessel invasion is a major prognostic factor in resected non-small cell lung cancer. Ann Thorac Surg. 1996; 62:1489–1493. [PubMed: 8893589]

30. Sanchez-Salcedo P, Berto J, de-Torres JP, Campo A, Alcaide AB, Bastarrika G, Pueyo JC, Villanueva A, Echeveste JI, Lozano MD, Garcia-Velloso MJ, Seijo LM, Garcia J, Torre W, Pajares MJ, Pio R, Montuenga LM, Zulueta JJ. Lung cancer screening: fourteen year experience of the Pamplona early detection program (P-IELCAP). Arch Bronconeumol. 2015; 51:169–176. [PubMed: 25641356]

31. Tammemagi CM, Pinsky PF, Caporaso NE, Kvale PA, Hocking WG, Church TR, Riley TL, Commins J, Oken MM, Berg CD. Lung cancer risk prediction: prostate, lung, colorectal and ovarian cancer screening trial models and validation. J Natl Cancer Inst. 2011; 103:1058–1068. [PubMed: 21606442]

32. Mizuno S, Takiguchi Y, Fujikawa A, Motoori K, Tada Y, Kurosu K, Sekine Y, Yanagawa N, Hiroshima K, Muraoka K. Chronic obstructive pulmonary disease and interstitial lung disease in patients with lung cancer. Respirology. 2009; 14:377–383. [PubMed: 19192220]

33. Wille MM, Thomsen LH, Petersen J, de Bruijne M, Dirksen A, Pedersen JH, Shaker SB. Visual assessment of early emphysema and interstitial abnormalities on CT is useful in lung cancer risk analysis. Eur Radiol. 2015:1–8.

34. Bishawi M, Moore W, Bilfinger T. Severity of emphysema predicts location of lung cancer and 5-y survival of patients with stage I non–small cell lung cancer. J Surg Res. 2013; 184:1–5. [PubMed: 23810745]

35. Henschke CI, Yip R, Boffetta P, Markowitz S, Miller A, Hanaoka T, Wu N, Zulueta JJ, Yankelevitz DF. Investigators I-E. CT screening for lung cancer: Importance of emphysema for never smokers and smokers. Lung Cancer. 2015; 88:42–47. [PubMed: 25698134]

**At a Glance Commentary**

**Scientific Knowledge on the Subject**

Screening for lung cancer with low-dose computed tomography (LDCT) is associated with a significant reduction in lung cancer-related mortality. Despite the clinical benefit of this modality, the high false positive rate is a substantial limitation. A reliable predictor of nodule malignancy is needed.

**What This Study Adds to the Field**

We found that the performance of semantic features was comparable to lung-RADS for subsequent scan in differentiating benign and malignant nodules, and outperformed lung-RADS for baseline scan in predicting subsequent occurrence of tumor. The predictive power of lung-RADS could be improved by incorporating additional semantic features. The semantic features would be helpful in assessing lung cancer risk in a screening cohort.
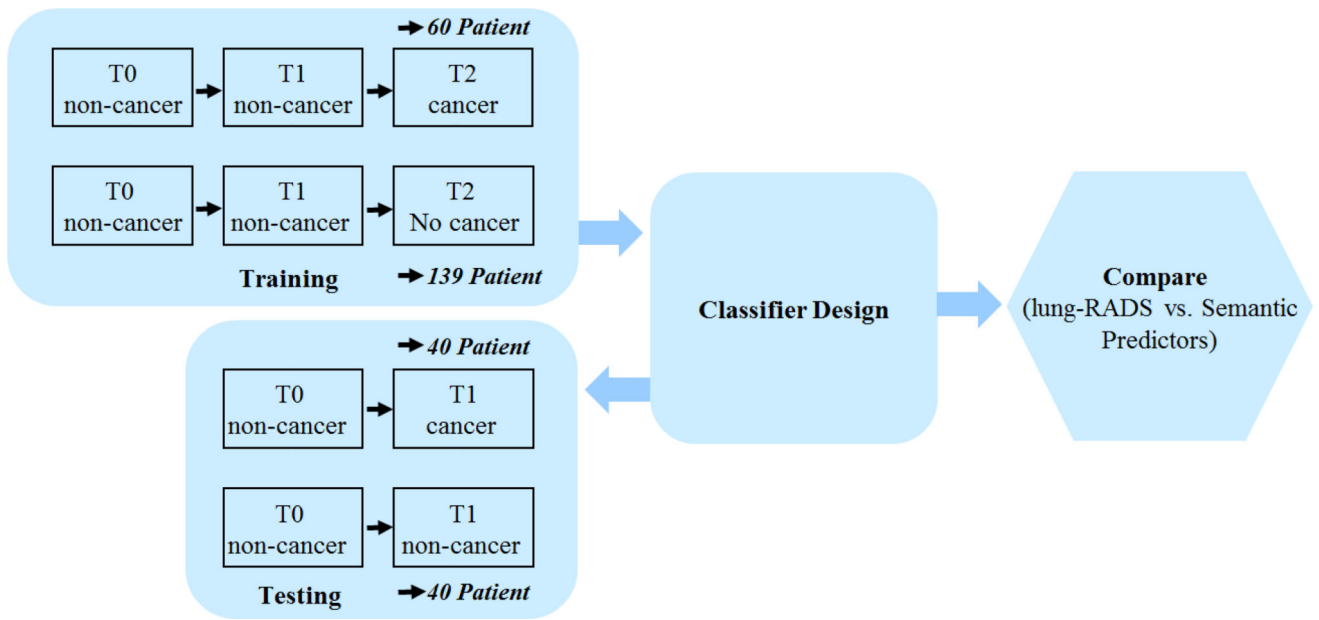
**Figure 1.**
Schema of the study
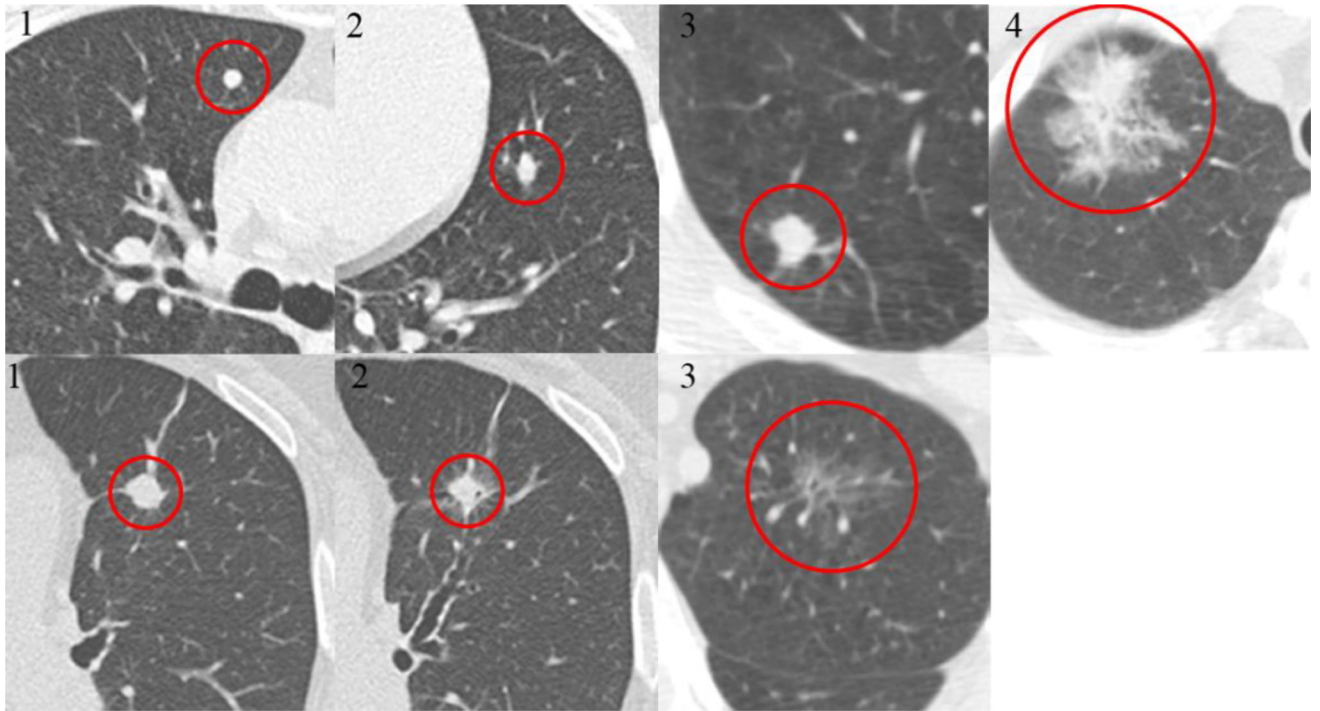
**Figure 2.**
Examples of scoring of radiological features (contour and border definition)
The top line is the examples of contour (1 means round, 2 means oval, 3 means irregular and 4 means extremely irregular), and the bottom line is the examples of border definition (1 means clear margin, 3 means poorly defined margin, and 2 means everything between them).

**Figure 3.**
Performance comparisons between semantic models and lung-RADS at three screening rounds (a: baseline screening, b: first follow-up scan, c: second follow-up scan)

**Figure 4.**
Nodules that smaller than 4mm at the baseline screening (T0) and developed tumor at the third screening round (T2)
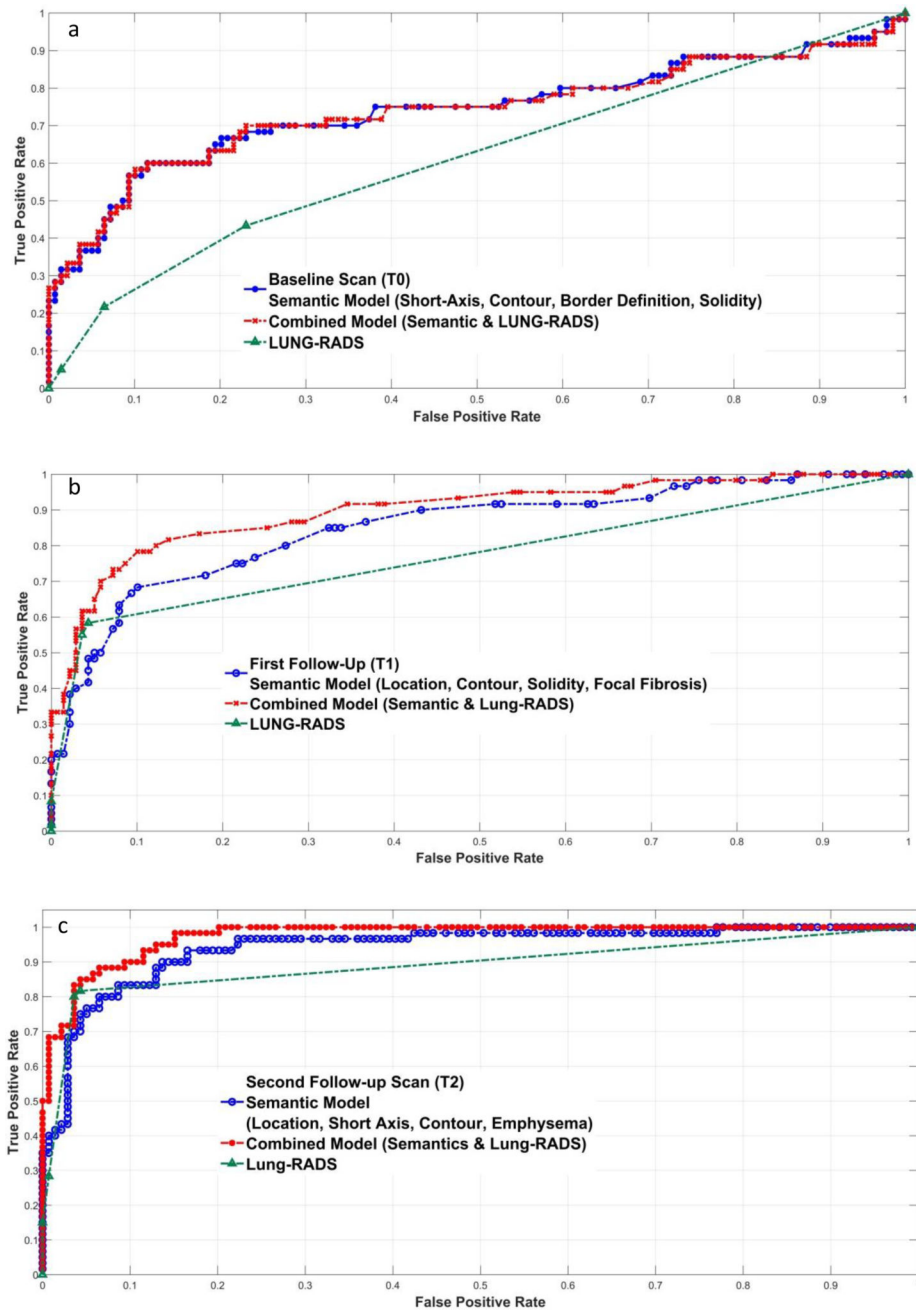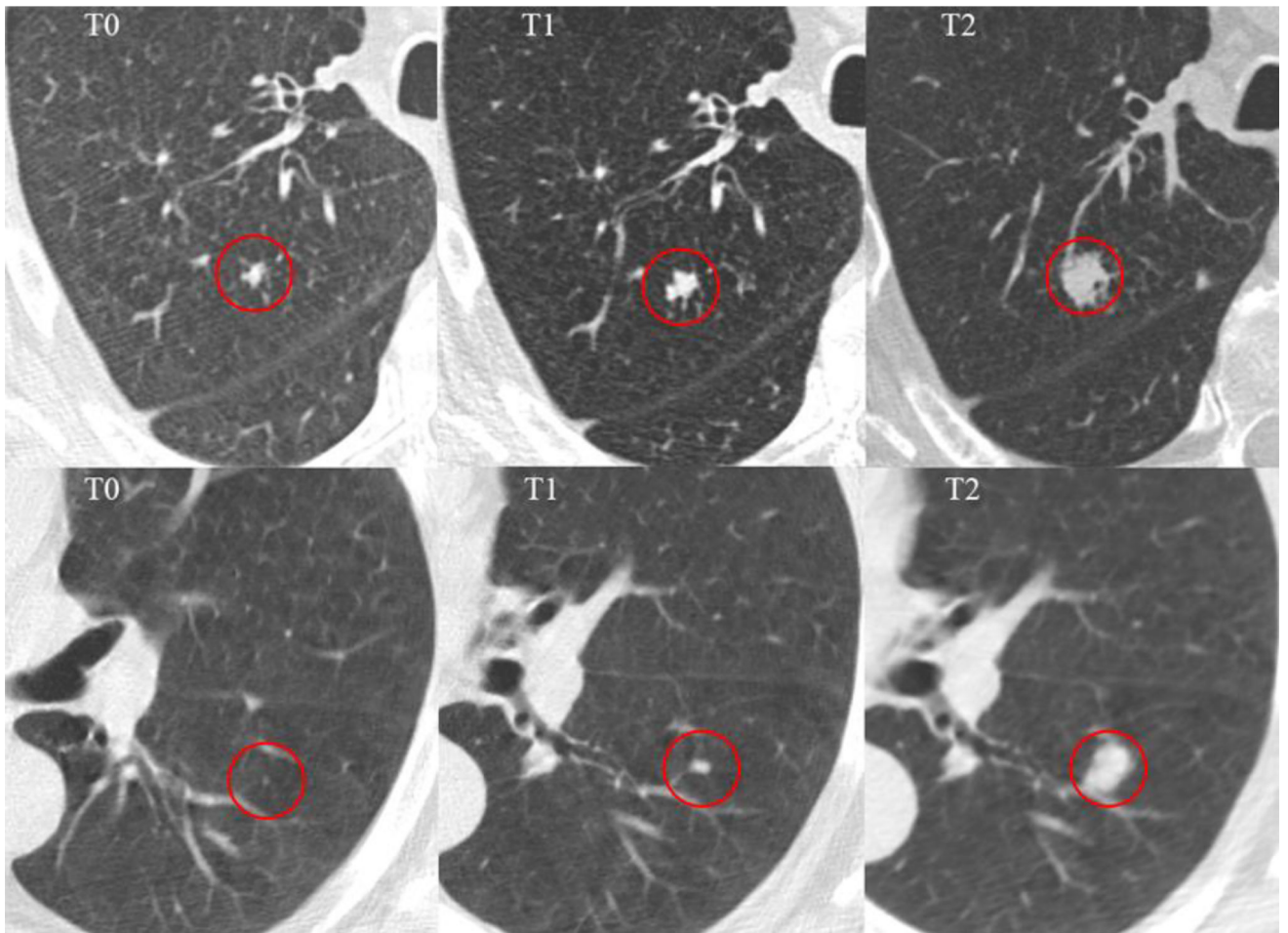
**Table 1**

Demographic characteristic of NLST subjects

| Characteristic | | | Training (N = 199) | | Testing (N = 80) | |
|---|---|---|---|---|---|---|
| | | | Lung Cancer Cases (N = 60) | nodule-positive controls (N = 139) | Lung Cancer Cases (N = 20) | nodule-positive controls (N = 20) |
| Age | | mean (SD) | 62.3 (4.9) | 63.1 (5.0) | 65.3 (5.2) | 65.2 (5.3) |
| Gender, N (%) | | Male | 29 (48.3) | 79 (56.8) | 19 (47.5) | 30 (75.0) |
| | | Female | 31 (51.7) | 60 (43.2) | 21 (52.5) | 10 (25.0) |
| Race, N (%) | | White | 58 (96.7) | 134 (96.4) | 39 (97.5) | 39 (97.5) |
| | | Other | 2 (3.3) | 5 (3.6) | 1 (2.5) | 1 (2.5) |
| Ethnicity, N (%) | | Hispanic or Latino | 1 (1.7) | 0 | 0 | 2 (5.0) |
| | | Neither Hispanic/Latino and Unknown | 59 (98.3) | 139 (100) | 40 (100) | 38 (95.0) |
| Current Smoker, N (%) | | Yes | 29 (48.3) | 76 (54.7) | 24 (60.0) | 20 (50.0) |
| | | No | 31(51.7) | 63 (45.3) | 16 (40.0) | 20 (50.0) |
| Pack-Year Smoking, mean (SD) | | Current smokers | 62.1 (18.8) | 63.0 (19.6) | 59.8 (16.5) | 66.6 (30.9) |
| | | Former smokers | 62.3 (29.7) | 58.0 (23.1) | 72.3 (31.3) | 82.1 (34.4) |
| Self-Reported History of COPD, N (%) | | Yes | 5 (8.3) | 12 (8.6) | 3 (7.5) | 4 (10.0) |
| | | No | 55 (91.7) | 127 (91.4) | 37 (92.5) | 36 (90.0) |
| Family history of lung cancer, N (%) | | Yes | 14 (23.3) | 23 (16.5) | 10 (25.0) | 9 (22.5) |
| | | No | 46 (76.7) | 116 (83.5) | 30 (75.0) | 31 (77.5) |
| Stage, N (%) | I | IA | 35 (58.3) | -- | 23 (57.5) | -- |
| | | IB | 11 (18.3) | -- | 6 (15) | -- |
| | II | | 2 (3.3) | -- | 5 (12.5) | -- |
| | III | | 7 (11.7) | -- | 4 (10.0) | -- |
| | IV | | 2 (3.3) | -- | 2 (5.0) | -- |
| | other, Unknown | | 3 (5.0) | -- | | -- |
| Histology, N (%) | | Adenocarcinoma | 41 (68.4) | -- | 28 (70.0) | -- |
| | | Squamous cell carcinoma | 5 (8.3) | -- | 3(7.5) | -- |
| | | Other, NOS, Unknown | 14 (23.3) | -- | 9 (22.5) | -- |

**Abbreviations:** COPD = chronic obstructive pulmonary disease.

**Table 2**

The performance of top semantic features, feature combinations, and lung-RADS in predicting nodule malignancy at each screening round (a: baseline scan, b: first follow-up scan, c: second follow-up scan)

### 2a Baseline scan (T0)

| Features | Semantics | | | Lung-RADS | | | Combined[†] |
|---|---|---|---|---|---|---|---|
| | Sensitivity/ Specificity | E[AUC]* [CI] | Testing AUC | Sensitivity/ Specificity | E[AUC] CI | AUC [CI] |
| Border definition | 0.579 / 0.834 | 0.719 [0.582,0.825] | 0.787 | 0.399 – 0.466 / 0.750 – 0.794 | 0.600 [0.404,0.719] | 0.743 [0.601,0.878] |
| Contour | 0.372 / 0.915 | 0.689 [0.484,0.864] | 0.835 | | | 0.685 [0.471,0.851] |
| Border definitionFocal fibrosis | 0.571 / 0.839 | 0.692 [0.542,0.825] | 0.789 | | | 0.732 [0.587,0.881] |
| Border definition solidity Vessel attachment | 0.549 / 0.858 | 0.731 [0.582,0.896] | 0.782 | | | 0.704 [0.512,0.913] |
| Short axial diameter Contour Border definition solidity | 0.525 / 0.888 | 0.741 [0.574,0.855] | 0.927 | | | 0.732 [0.555,0.852] |

### 2b First follow-up scan (T1)

| Features | Semantics | | | Lung-RADS | | | Combined[†] |
|---|---|---|---|---|---|---|---|
| | Sensitivity/ Specificity | E[AUC]* [CI] | Testing[§] AUC | Sensitivity/ Specificity | E[AUC] CI | AUC [CI] |
| Contour | 0.689 / 0.898 | 0.823 [0.722,0.944] | 0.835 | 0.537 – 0.621/0.949 – 0.970 | 0.760 [0.614,0.919] | 0.875 [0.775,0.944] |
| Border definition | 0.695 / 0.815 | 0.741 [0.595,0.864] | 0.787 | | | 0.846 [0.752,0.983] |
| Location Contour | 0.678 / 0.905 | 0.833 [0.739,0.925] | 0.820 | | | 0.886 [0.793,0.969] |
| Contour Vessel attachment Focal emphysema | 0.687 / 0.901 | 0.882 [0.776,0.966] | 0.843 | | | 0.835 [0.525,0.967] |
| Location Contour solidity Focal fibrosis | 0.665 / 0.907 | 0.897 [0.807,0.974] | 0.825 | | | 0.897 [0.807,0.974] |

### 2c Second follow-up scan (T2)

| Features | Semantics | | | Lung-RADS | | | Combined[†] |
|---|---|---|---|---|---|---|---|
| | Sensitivity/ Specificity | E[AUC]* [CI] | Testing AUC | Sensitivity/ Specificity | E[AUC] CI | AUC [CI] |
| Contour | 0.829 / 0.871 | 0.876 [0.742,0.97] | 0.835 | 0.767 – 0.847/0.953 – 0.973 | 0.867 [0.760,0.973] | 0.958 [0.877,0.993] |
| Border definition | 0.729 / 0.806 | 0.802 [0.642,0.924] | 0.787 | | | 0.938 [0.845,0.992] |
| Location Contour | 0.816 / 0.881 | 0.897 [0.757,0.979] | 0.820 | | | 0.962 [0.883,0.991] |
| Contour Vessel attachment Focal emphysema | 0.823 / 0.881 | 0.903 [0.786,0.974] | 0.843 | | | 0.965 [0.894,0.994] |

**2c Second follow-up scan (T2)**

| Features | Semantics | | Testing AUC | Lung-RADS | | Combined† |
| --- | --- | --- | --- | --- | --- | --- |
| | Sensitivity/ Specificity | E[AUC]* [CI] | | Sensitivity/ Specificity | E[AUC] CI | AUC [CI] |
| Location Short axial diameter Contour Focal emphysema | 0.744 / 0.943 | 0.947 [0.881,0.987] | 0.923 | | | 0.968 [0.913,0.995] |

†
It indicates the combination of lung-RADS with semantic features.

§
there are only 2 screening rounds for the testing cohort, so we used the same dataset for baseline and the first follow-up scan.

*
E[AUC] means average area under the receiver-operating characteristic curve, CI stands for confidence interval.

**Table 3**

Agreement of semantic features between 2 readers

| CT features | Kappa | intra-class correlation of coefficient (ICC) |
|---|---|---|
| Location | 1 | |
| Fissure attachment | 1 | |
| Pleural attachment | 1 | |
| Contour | 0.71 | |
| Lobulation | 0.51 | |
| Concavity | 0.64 | |
| Border definition | 0.71 | |
| Spiculation | 0.69 | |
| solidity | 0.95 | |
| Air bronchogram | 1 | |
| Bubble-like lucency | 1 | |
| Vascular convergence | 0.48 | |
| Thickened adjacent bronchovascular bundles | 1 | |
| vessel attachment | 0.90 | |
| Pleural retraction | 1 | |
| Peripheral emphysema | 0.43 | |
| Peripheral fibrosis | 0.77 | |
| Nodules in primary tumor lobe | 1 | |
| Nodules in non-tumor lobes | 0.75 | |
| Lymphadenopathy | 0.64 | |
| Long axial diameter | | 0.940 (0.890 – 0.968) |
| Short axial diameter | | 0.960 (0.834 – 0.985) |