European Radiology (2019) 29:6100–6108
https://doi.org/10.1007/s00330-019-06213-9

CHEST

Comparison of prediction models with radiological semantic features and radiomics in lung cancer diagnosis of the pulmonary nodules: a case-control study

Wei Wu[1,2] · Larry A. Pierce[1] · Yuzheng Zhang[3] · Sudhakar N. J. Pipavath[1] · Timothy W. Randolph[4] · Kristin J. Lastwika[5,6] · Paul D. Lampe[5,6] · A. McGarry Houghton[4,6,7] · Haining Liu[2] · Liming Xia[1] · Paul E. Kinahan[2]

Received: 28 June 2018 / Revised: 1 March 2019 / Accepted: 2 April 2019 / Published online: 21 May 2019

Abstract

Purpose To compare the ability of radiological semantic and quantitative texture features in lung cancer diagnosis of pulmonary nodules.

Materials and methods A total of $N = 121$ subjects with confirmed non-small-cell lung cancer were matched with 117 controls based on age and gender. Radiological semantic and quantitative texture features were extracted from CT images with or without contrast enhancement. Three different models were compared using LASSO logistic regression: "CS" using clinical and semantic variables, "T" using texture features, and "CST" using clinical, semantic, and texture variables. For each model, we performed 100 trials of fivefold cross-validation and the average receiver operating curve was accessed. The AUC of the cross-validation study (AUC_{CV}) was calculated together with its 95% confidence interval.

Results The AUC_{CV} (and 95% confidence interval) for models T, CS, and CST was 0.85 (0.71–0.96), 0.88 (0.77–0.96), and 0.88 (0.77–0.97), respectively. After separating the data into two groups with or without contrast enhancement, the AUC (without cross-validation) of the model T was 0.86 both for images with and without contrast enhancement, suggesting that contrast enhancement did not impact the utility of texture analysis.

Conclusions The models with semantic and texture features provided cross-validated AUCs of 0.85–0.88 for classification of benign versus cancerous nodules, showing potential in aiding the management of patients.

Key Points

- *Pretest probability of cancer can aid and direct the physician in the diagnosis and management of pulmonary nodules in a cost-effective way.*
- *Semantic features (qualitative features reported by radiologists to characterize lung lesions) and radiomic (e.g., texture) features can be extracted from CT images.*

Paul E. Kinahan and Liming Xia have equal contribution as the corresponding authors.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00330-019-06213-9) contains supplementary material, which is available to authorized users.

✉ Liming Xia
xialiming2017@outlook.com

✉ Paul E. Kinahan
kinahan@uw.edu

[1] Department of Radiology, Tongji Hospital, Tongji Medical College affiliated to Huazhong University of Science and Technology, 1095 Jiefang Ave, Wuhan, Hubei 430000, People's Republic of China

[2] Department of Radiology, University of Washington, 1959 NE Pacific St, Seattle, WA 98105, USA

[3] Program in Biostatistics and Biomathematics, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

[4] Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

[5] Translational Research Program, Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

[6] Human Biology Divisions, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

[7] Division of Pulmonary and Critical Care, University of Washington Medical Center, Seattle, WA, USA

• *Input of these variables into a model can generate a pretest likelihood of cancer to aid clinical decision and management of pulmonary nodules.*

Keywords Lung cancer · Tomography · Radiomics · Semantics · Statistical models

Abbreviations

CI	Confidence intervals
CT	X-ray computed tomography
NSCLC	Non-small cell lung cancer
VM	The bounding volume maximum length
VOI	Volume of interest

Introduction

Management of pulmonary nodules is a problem in clinical scenarios, in part due to increasing use of multislice computed tomography (CT) with contiguous thin sections, considered the gold standard for pulmonary nodule detection [1]. For both screening and incidental findings, it can be challenging to classify pulmonary nodules between benign and malignant at first presentation, and overdiagnosis is a challenging issue [2–4].

It has been accepted by clinicians that the pretest probability of cancer can aid and direct the physician in the diagnosis of pulmonary nodules in a cost-effective way. While many clinicians and radiologists rely on clinical experience to assess the likelihood that a nodule is cancerous, there is recent research using quantitative methods to inform the assessment of a newly discovered pulmonary nodule. Prediction models have been developed that utilize both clinical and radiological data to classify nodules. A few of these models [5–7] have been independently validated to exhibit classification accuracies (as measured by the area under the receiver operator characteristic curve, or AUC) in the range of 0.57 to 0.90 [8–10].

The emerging field of "radiomics" has been recently studied as a method to extract more quantitative information from radiological images to aid clinical decisions. In radiomics, mathematical formulae that quantitatively describe various image qualities (e.g., contrast, homogeneity, texture) are defined in a region and compared to outcomes to form a predictive model. Providing a definition of the radiomics paradigm, Gilles et al stated that "Images Are More than Pictures, They Are Data" [11]. Specifically, radiology images are transformed to higher dimensional data to search for useful correlations that define a useful radiographic phenotype before, during, or after therapy [12]. Recent developments in radiomics indicate that quantitative analysis of images can help discriminate different types of cancer, predict disease outcomes, and determine response to treatment [13–17]. The application of radiomics methods to the management of pulmonary nodules has been suggested [18] and evaluated in patients in at least two previous studies [19, 20].

Quantitative texture features may compliment traditional prediction models for pulmonary nodule diagnosis, and a model that integrates these approaches may lead to increased predictive classification accuracy. In this study, we created individual models with exclusive texture or clinical and semantic features as well as their combination in order to assess the predictive power of each. We also examined the ability of quantitative texture features to compliment clinical data (e.g., smoking history) and semantic imaging features (e.g., spiculated nodule, ground glass opacity), with the goal of determining how radiomics may help with pulmonary nodule classification.

Materials and methods

Patients

Data were prospectively collected from October 2010 to September 2015 for a cohort of 492 subjects enrolled in the Fred Hutch Lung Cancer Early Detection and Prevention Clinic (LCEDC) under an active institutional review board, and written informed consents were acquired. All of the cases were reviewed by a multidisciplinary lung nodule board, and CT images were collected at 3–12-month intervals until they were determined to be malignant or benign. For our retrospective study of the LCEDC data, IRB approval and a waiver of written informed consent were obtained. Inclusion criteria were as follows: (1) Only the patients with at least one pulmonary nodule that was at least 1 mm in diameter, when using lung parenchymal CT display thresholds, on the baseline CT image were included [7]. (2) For patients with malignant tumors, only subjects with nodules that were determined to be non-small cell lung cancer (NSCLC) by histopathology before the consent date of May 2014 were included. (3) For benign nodules, only those confirmed by histopathologic examination of tissue obtained via surgical resection, or the lesion was found to be stable radiographically for at least 2 years of follow-up, or resolved under CT surveillance, were included. There were $N = 125$ NSCLC case subjects included, which were 1:1 matched with subjects with benign nodules. Matching was first based on gender (exact match), then age (±2 years). In attempting to match the NSCLC cases with controls for pack years of smoking, we found that matching was not possible due to the limited sample size. Clinical demographics of age, gender, race, BMI, smoking behavior, including current smoking status, pack-years, and histology

were recorded. We excluded 12 subjects who did not have qualifying CT scan or did not have full clinical demographic records. Figure 1 summarizes the flowchart of the study enrolling process and reasons for patient exclusion. A total of 238 patients were included in this study.

CT imaging acquisition

The CT scans used in this study were from multiple institutions using local imaging protocols. This resulted in non-uniform imaging protocols across patients in the study. The CT images were acquired with or without contrast enhancement. Details regarding the different protocols and acquisition parameters are in the Supplementary Material.

Nodule segmentation

For each patient, only the largest nodule confirmed by histopathology, or confirmed to be benign by surveillance, was chosen for analysis. Contours defining the nodule region (Fig. 2) were manually delineated slice-by-slice on the transaxial CT images by a thoracic radiologist (W.W., 6 years of experience) using the MIM image viewing and analysis software (MIM Software, Inc.). For the nodules which are larger than 5 mm in long-axis diameter, the slice thickness of the images for contouring is 2.5 mm, while for nodules which are equal or less than 5 mm, the slice thickness for contouring is 1.25 mm. Lesion volumes were first contoured in a mediastinal window (WW 350 HU WL 40 HU) to identify the boundaries with the chest wall or other soft tissues, then in a lung window (WW 1324 HU WL -362 HU) to capture the extent in the lung parenchyma. A 3D volume of interest (VOI) was created to encompass the entire nodule and then converted to a binary mask.

Radiomics feature extraction

For each patient, the CT image and corresponding binary VOI mask were imported into the PORTS radiomics software package [21] (details in Supplementary Material). PORTS computed 42 quantitative image and texture metrics from the voxels within the VOI (Supplementary Table S2 and Fig. 3).

Semantic feature extraction

A second thoracic radiologist (S.P., 12 years of experience), blinded to clinical and histologic findings, evaluated semantic features including nodule count, location, size, shape, density, and the visual presence and the types of invasiveness, lymphadenopathy, emphysema, fibrosis, and asbestosis. In total, 13 categories of semantic features from CT images were extracted for the analysis.

Statistical analysis

Univariate tests were performed to assess the difference of the estimated features between the case and control status using the PASW Statistics16.0™ analysis package (SPSS Inc.). The Fisher exact test was used for the categorical data and the Mann-Whitney *U* test was used for continuous data.

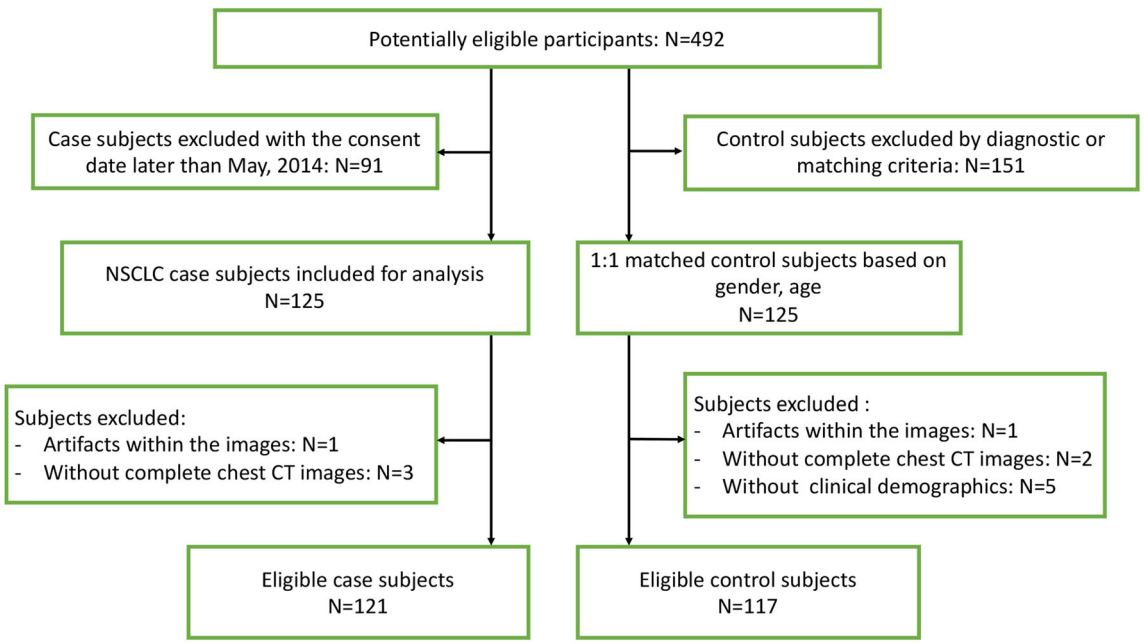

Fig. 1 Flowchart of the study enrolling process and reasons for patient exclusion

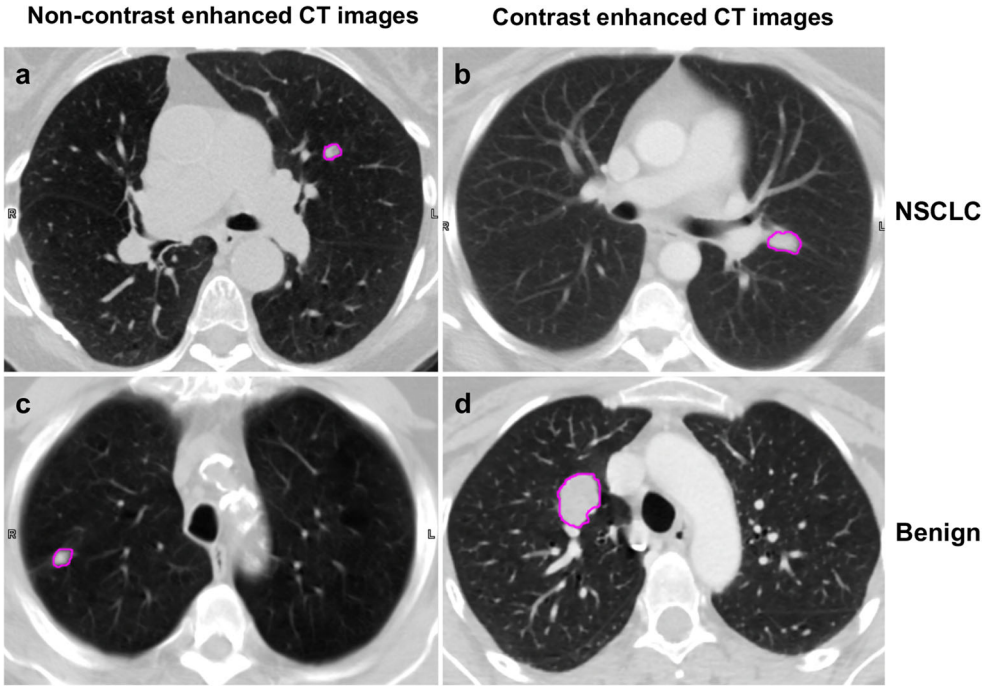

Fig. 2 Example images illustrating the categories of patient (NSCLC vs. benign, and contrast vs. non-contrast imaging protocol) and pulmonary nodule segmentation of each

LASSO penalized logistic regression [22, 23] was used as a variable-selection procedure to build a low-dimensional logistic regression model. LASSO was performed using the glmnet package in R version 3.4.3 [24]. We constructed three models based on subsets of variables (details in Supplementary Material). The three models are "CS" using the clinical and semantic variables only, "T" using the texture features as well as the CT scan in kilovoltage peak (kVp), and "CST" using all of the clinical, semantic, and texture variables. In each case, the variables selected by the LASSO were used to fit an ordinary logistic regression model and estimate the regression coefficients. For the resulting models, a receiver operator characteristic curve (ROC) was created and the area under the curve (AUC) was computed to evaluate the prediction

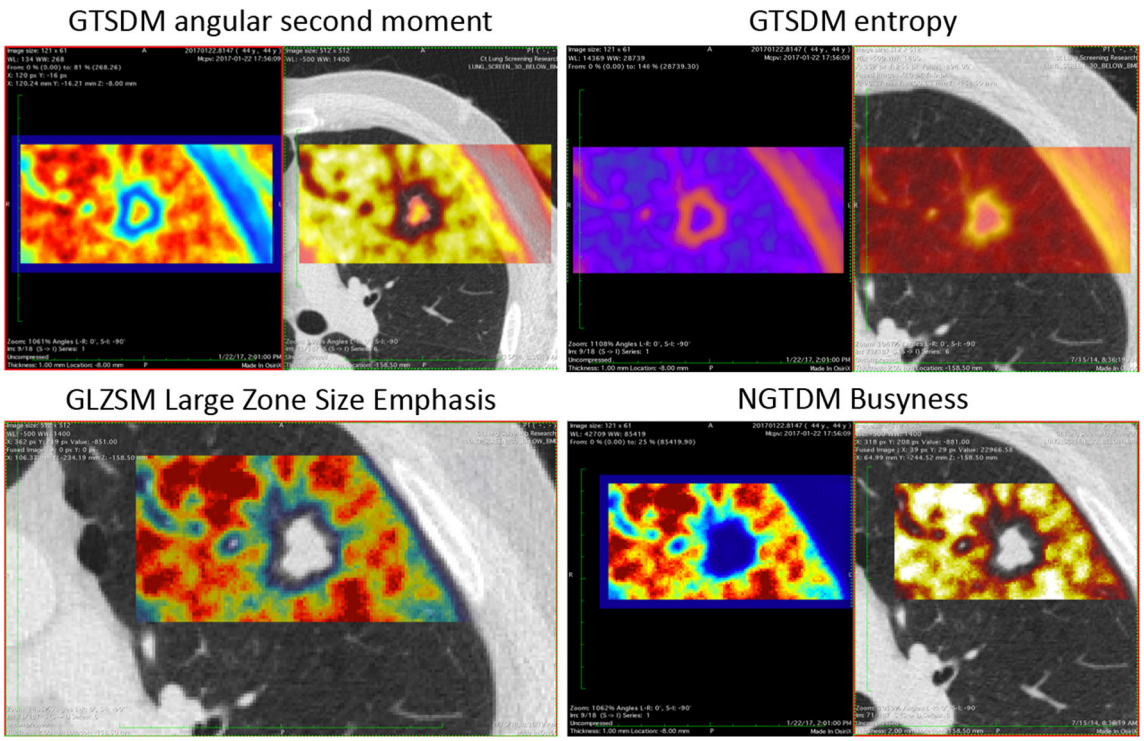

Fig. 3 Examples of PORTS texture analysis showing the quantitative values of four different texture features

capability for each model separately. For each model, we performed 100 trials of fivefold cross-validation, where four-fifths of data was used as training and one-fifth of the data was used for testing. The AUC of the cross-validation study (AUC_{CV}) was calculated together with its 95% confidence interval.

Since the CT images were approximately equally divided between those with and without contrast enhancement, we also assessed the impact of contrast enhancement on texture features within the CT images by comparing the abilities of the texture-only models (i.e., the model T) to predict nodule status. The goal was to evaluate if there was any impact of contrast enhancement on the texture model and no attempt was made to compare with the CS and CST models described above.

Results

Clinical characteristics

Clinical characteristics for the total 238 patients and their distribution in case or control group are summarized in Table 1. The mean age was 63.8 ± 10.6 years. There were 121 NSCLC subjects and 117 control subjects which included 64 cases with the nodules stable for at least 2 years, 24 cases with the nodules resolved during the follow-up, and 29 cases with confirmed histopathology diagnosis. There was no difference for the mean age, proportion of males, or BMI between the NSCLC and control groups. However, the control and NSCLC groups exhibited some difference in smoking history with the control group containing patients who were former (i.e., not current) smokers or who had never smoked, while there were more current smokers in the NSCLC group ($p < 0.05$). We further reviewed all cases of former smokers and found that the minimal cessation time was 18 days. For most of the former smokers, the cessation time was longer than 1 year and just 5 cases (5/123) had a cessation time of less than 1 year. The control group had a mean of 25.3 pack-years smoked versus 32.4 pack-years for the NSCLC group; however, there was no significant difference ($p = 0.064$). We note that reported univariate p values in Table 1 (and Supplementary Table S1) are not claims about the statistical significance of the models, as it relates to the population of all cancer samples, but rather they are reported here for descriptive and comparative purposes only.

Semantic and texture features

The Fisher exact test and Mann-Whitney U test revealed that six categories of semantic features have the potential for distinguishing between cases and controls. (Supplementary

Table 1 Clinical characteristics of participants with NSCLC and benign nodules

Category: data type	Full sample	Benign	NSCLC	p* value
Age: mean (± SD)	63.8 (±10.6)	63.0 (±11.5)	64.4 (±9.52)	0.42
Male: N (%)	106 (44.5)	53 (45.3)	53 (43.8)	0.90
BMI: mean (± SD)	27.8 (±7.2)	27.7 (±7.3)	28.0 (±7.1)	0.49
Pack-years: mean (± SD)	28.9 (±26.0)	25.3 (±22.1)	32.4 (±29.0)	0.06
Smoking history: N (%)				0.04
Never smoker	43 (18.1)	26 (22.2)	17 (14.0)	
Former smoker	123 (51.7)	64 (54.7)	59 (48.8)	
Current smoker	72 (30.3)	27 (23.1)	45 (37.2)	
Histology: N (%)				–
Adenocarcinoma	89 (37.4)	–	89 (73.6)	–
Adenocarcinoma in situ	13 (5.5)	–	13 (10.7)	–
Squamous	17 (7.1)	–	17 (14.0)	–
Large cell	2 (0.8)	–	2 (1.7)	–
Infections	10 (4.2)	10 (8.5)	–	–
Inflammation	17 (7.1)	17 (14.5)	–	–
Choroid hamartoma	1 (0.4)	1 (0.9)	–	–
Fibrosis/scarring	1 (0.4)	1 (0.9)	–	–

p values are uncorrected for multiple comparisons and are used only descriptively

A former smoker is defined as an adult who has smoked at least 100 cigarettes in his or her lifetime but who had quit smoking at the time of interview. This is consistent with the definition of the National Health Interview Survey (NHIS) from Centers for Disease Control and Prevention (CDC)

*Mann-Whitney U test for continuous data, Fisher exact test for categorical data

Table S1). We report here the unadjusted p values as a means to describe their relative strength for discrimination. These features include emphysema degree ($p = 0.01$), nodule location ($p = 0.02$), size including long-axis diameters (L), short-axis diameters (S), height (H), and the bounding volume maximum length (VM, defined as VM=$\sqrt{L^2 + S^2 + H^2}$) ($p < 0.001$ for each), shape ($p < 0.001$), invasiveness ($p < 0.001$), and lymphadenopathy ($p < 0.001$). No difference was detected between two groups for pattern of emphysema, region of emphysema, fibrosis, nodule count, or nodule density. The distribution of each semantic feature for the two groups is summarized in Supplementary Table S1. With the exception of features F9, F16, F17, and F31, all quantitative texture metrics showed evidence of a difference between the NSCLC and control groups ($p < 0.05$).

Multivariate and ROC analysis of clinical, semantic, and texture features in classification of cases versus controls

Model T used the texture features and kVp as input variables. The LASSO regression selected 6 (out of 43) variables: F10 (GTSDM-Correlation), F19 (GTSDM-Information Correlation1), F21 (GTSDM-Autocorrelation), F22 (GTSDM-Dissimilarity), F41 (GLZSM-Zone Size Non-Uniformity), and F42 (GLZSM-Zone Size Percentage). The AUC_{CV} for model T was 0.85 with a 95% confidence interval of 0.71–0.96. Model CS used the clinical and semantic variables only. The LASSO selected 5 (out of 22) variables: smoothness, spiculation, invasiveness, lymphadenopathy, and VM. The AUC_{CV} was 0.88 with a 95% confidence interval of 0.77–0.96. Model CST combined all of the clinical, semantic, and texture variables. The LASSO selected 7 (out of 64) variables: smoothness, spiculation, invasiveness, lymphadenopathy, VM, F19 (GTSDM-Information Correlation1), and F22 (GTSDM-Dissimilarity). The AUC_{CV} was 0.88 with a 95% confidence interval of 0.77–0.97. These results are presented in Table 2 and Fig. 4.

Comparison of models of contrast versus non-contrast images

There were 125 patients with contrast-enhanced images and 113 patients with non-contrast-enhanced images. The AUC (without cross-validation) of the model T was 0.86 both for images with and without contrast enhancement. When these data are combined into a single population, the AUC was also 0.86. The three ROC curves and AUCs are shown in Supplementary Fig. S1.

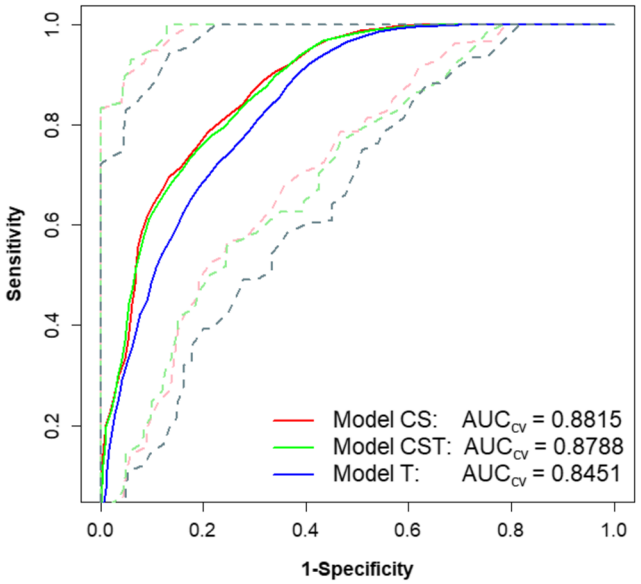

Fig. 4 ROCs and cross-validated AUC_{CV} of three different models. Solid lines are the fivefold cross-validated AUCs and the dotted lines in corresponding colors refer to the 95% confidence interval for each AUC_{CV}

Discussion and conclusion

We examined the capabilities of clinical, semantic, and image texture features to assess cancer risk in pulmonary nodules. Our cross-validated multivariate analysis shows that models CS and CST had $AUC_{CV} = 0.88$, slightly larger than model T with $AUC_{CV} = 0.85$ (Fig. 4). The 95% confidence intervals of all three models show that the ROC performance largely overlapped, demonstrating similar performance of all three models. While the AUC is a global measure accounting for all possible sensitivity and specificity thresholds, for comparison purposes, we can estimate that for a true-positive rate (sensitivity) of 90%, the false-positive rate (1-specificity) ranges from 34 to 38%. If the true-positive rate is relaxed to 80%, the false-positive rate is reduced to a range of 23 to 31% for the three methods (Fig. 4). These models were controlled

Table 2 Prediction performance of the three different models

Model	Subjects	Variables considered	Number of LASSO-selected top predictor variables	AUC (all data)	AUC_{CV} (95% CI)
T	238	43	6	0.86	0.85 (0.71–0.96)
CS	238	22	5	0.89	0.88 (0.77–0.96)
CST	238	64	7	0.89	0.88 (0.77–0.97)

for gender and age, but not smoking history or smoking pack-years. To check the impact of these factors we separately included smoking history and smoking pack-years into the models. When adjusted for smoking history, the AUC_{CV} values were 0.88, 0.88, and 0.85 for the CS, CST, and T models. In other words, there was no detectable change compared to the models without any adjustment. When adjusted for smoking pack-years, the AUC_{CV} values increased very slightly to 0.89, 0.89, and 0.86 for the CS, CST, and T models, respectively—very similar to the performance of the models with the smoking history adjustment. The similar results may be due to a high correlation between smoking pack-years and smoking history.

These results indicate that models based on semantic features determined by an experienced thoracic radiologist will likely slightly outperform models based only on computed texture features. This is supported by the recent work of Liu et al who demonstrated that a set of clinically relevant radiologic features can be readily scored to effectively determine lung cancer risk for solitary pulmonary nodules [25].

Even so, the texture features perform nearly as well and can be used to standardize performance across clinical sites. In addition, work continues in the exploration of radiomics methods with larger cohorts so further performance improvements are also possible.

A key component is the integration of accurate and efficient lung nodule delineation or segmentation methods. Evaluation of this component was beyond the scope of this study, but we note that manual delineation is time-consuming and known to suffer from intra- and inter-reader variability [26]. However, recent studies have shown that some automated and semi-automated lung nodule segmentation tools can reduce variability [27]. In the event an automated or semi-automated lung nodule delineation is incorporated, the radiomics workflow could be very efficient. This might be particularly helpful in instances where radiologists have less lung specific training or expertise.

To our knowledge, the inclusion of quantitative texture features with semantic features in our predictive model is unique among similar studies, including the Mayo Clinic model [6], the Veteran's Affairs model [5], the Herder model [28], and the model from Brock University [7], which use clinical and limited semantic features. One 2016 study found that a model using texture features performed statistically similarly to that using the Brock University model on the NLST (National Lung Screening Trial) dataset with accuracy of 78% [19]. Another 2018 study of a relative small sample demonstrated a radiomics-only model with the ability to discriminate between malignant and benign pulmonary nodules with an accuracy of 84% [20].

To evaluate if it is reasonable to include both non-contrast-enhanced and contrast-enhanced images in our model, we performed a substudy to compare the texture feature performance in the contrast and non-contrast groups. Our analysis showed that quantitative texture features perform similarly in patient populations with and without contrast-enhanced images (AUC of 0.86). Although one previous study reported that the predictive performance of quantitative texture features is superior in non-contrast-enhanced images (AUC = 0.86) than those with contrast (AUC = 0.83) [29], a more recent study on discrimination of lung-invasive adenocarcinoma validated the radiomics signature in two independent non-contrast-enhanced and contrast-enhanced cohorts and found the performance was similar [30], consistent with our findings. In the discussion for this study, it was hypothesized that the radiomics signature may be independent of the injection of a contrast agent as the selected radiomics features measure the correlation, uniformity, or deviation of the pixels, which reduces the effect of contrast increase of the overall image intensity. Although plausible, this hypothesis should be confirmed. Regardless, as our model is independent with regard to the use of contrast and its combination of images with multiprotocols, this may allow for broader applicability in clinical care.

There were several limitations in our study. First, this was a retrospective analysis with a relatively small population that included prediagnostic (screening) and diagnostic patients. Second, we did not have an independent patient population as a validation cohort, as would be required to assess the accuracy the models. Instead, these models provide a proof of principle regarding the role of radiomic and semantic features to facilitate the clinical assessment of pulmonary nodules. For validation, a follow-up study with an independent patient population is needed. Potentially another limitation is the non-standardized image acquisition and definition of radiological semantic features. There is an unresolved debate on the requirement for standardized protocols versus the need for methods that work in a typical imaging network [31]. In general, we expect standardized acquisition and processing to enable distinguishing smaller differences between tumor phenotypes. In clinical practice, however, a wide variety of scanning techniques and parameters are used. In CT phantom studies, differences in slice thickness and reconstruction algorithm (standard versus lung) were shown to substantially change radiomic feature values [32]. However, recent studies on the reproducibility of features extracted from lung CT images have demonstrated that a subset of radiomic features are reproducible and informative even in the presence of scan protocol differences [33, 34]. In addition, Zhao et al [33] and Lu et al [35] showed that use of both 1.25-mm and 2.5-mm slice thicknesses is acceptable if the use of "lung" image reconstruction filters is avoided, as was the case for our study. Finally, we note that a recent study concluded that voxel-size resampling is an appropriate preprocessing step for image datasets acquired with variable voxel sizes to obtain more reproducible CT features [36]. As noted in the supplementary

data, this was the procedure used for the radiomics analysis presented here. Another potential limitation of our approach became apparent for 8 of the 117 control cases that had nodules smaller than 4 mm. Since nodules less than 4 to 6 mm are considered less significant clinically [1, 37], and the small size may prevent calculation of texture metrics, the impact of these smaller lesions is unclear. In our analysis, only one texture metric was not calculable for 7 of the cases with nodules smaller than 4 mm. We estimated these missing values using the *k*-nearest-neighbor approach by grouping subjects with the similar characteristics of age, race, smoking history, and BMI status [38]. When the most important texture variables were chosen using LASSO regression as described above, the non-calculable texture metric was not selected. Thus, for our study, we believe there was no impact from these few small nodules. However, the issue of non-calculable texture metrics for very small nodules should be considered in future studies. If there is an observable effect, then the use of interpolation to smaller voxels might be considered.

Despite these limitations, we conclude that using semantic and quantitative texture features from CT images has the potential to aid and facilitate the clinical decision or management of pulmonary nodules.

Acknowledgments At the University of Washington Medical Center, we thank Steven R. Bowen, PhD, for helpful suggestions on the project, Nina A. Mayr, MD, and William T. Yuh for providing the access to MIM software and guidance on how to use the software.

Funding This study has received funding by NIH grants U01CA148131, U01185097, U01186157, P30CA015704, and F32CA200265, as well as National Natural Science Foundation of China (No. 81471637).

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Paul E. Kinahan.

Conflict of interest Paul E. Kinahan received a research grant from GE Healthcare outside of this work, and is the cofounder of PET/X LLC.

All other coauthors 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 Timothy W. Randolph and Yuzheng Zhang (two coauthors in our paper) have significant statistical expertise.

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

Ethical approval Institutional Review Board approval was obtained.

Methodology
- retrospective
- case-control study/diagnostic or prognostic study
- multicenter study

References

1. 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 284:228–243
2. Carter SM, Barratt A (2017) What is overdiagnosis and why should we take it seriously in cancer screening? Public Health Res Pract 27:e2731722
3. Edey AJ, Hansell DM (2009) Incidentally detected small pulmonary nodules on CT. Clin Radiol 64:872–884
4. Dziedzic R, Rzyman W (2014) Incidentally diagnosed pulmonary nodules: a diagnostic algorithm. Kardiochir Torakochirurgia Pol 11: 397–403
5. Gould MK, Ananth L, Barnett PG (2007) A clinical model to estimate the pretest probability of lung cancer in patients with solitary pulmonary nodules. Chest 131:383–388
6. Swensen SJ, Silverstein MD, Ilstrup DM, Schleck CD, Edell ES (2008) The probability of malignancy in solitary pulmonary nodules. Arch Intern Med 157:849–855
7. McWilliams A, Tammemagi MC, Mayo JR et al (2013) Probability of cancer in pulmonary nodules detected on first screening CT. N Engl J Med 369:910–919
8. Schultz EM, Sanders GD, Trotter PR et al (2008) Validation of two models to estimate the probability of malignancy in patients with solitary pulmonary nodules. Thorax 63:335–341
9. Al-Ameri A, Malhotra P, Thygesen H et al (2015) Risk of malignancy in pulmonary nodules: a validation study of four prediction models. Lung Cancer 89:27–30
10. Talwar A, Rahman NM, Kadir T, Pickup LC, Gleeson F (2017) A retrospective validation study of three models to estimate the probability of malignancy in patients with small pulmonary nodules from a tertiary oncology follow-up center. Clin Radiol 72:177.e1–177.e8
11. Gillies RJ, Kinahan PE, Hricak H (2016) Radiomics: images are more than pictures, they are data. Radiology 278:563–577
12. van Griethuysen JJM, Fedorov A, Parmar C et al (2017) Computational radiomics system to decode the radiographic phenotype. Cancer Res 77:e104–e107
13. Huang Y, Liu Z, He L et al (2016) Radiomics signature: a potential biomarker for the prediction of disease-free survival in early-stage (I or II) non—small cell lung cancer. Radiology 281:947–957
14. Coroller TP, Agrawal V, Huynh E et al (2017) Radiomic-based pathological response prediction from primary tumors and lymph nodes in NSCLC. J Thorac Oncol 12:467–476
15. Coroller TP, Grossmann P, Hou Y et al (2015) CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. Radiother Oncol 114:345–350
16. Huang YQ, Liang CH, He L et al (2016) Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. J Clin Oncol 34:2157–2164
17. Song SH, Park H, Lee G et al (2017) Imaging phenotyping using radiomics to predict micropapillary pattern within lung adenocarcinoma. J Thorac Oncol 12:624–632
18. Wilson R, Devaraj A (2017) Radiomics of pulmonary nodules and lung cancer. Transl Lung Cancer Res 6:86–91
19. Hawkins S, Wang H, Liu Y et al (2016) Predicting malignant nodules from screening CT scans. J Thorac Oncol 11:2120–2128
20. Chen CH, Chang CK, Tu CY et al (2018) Radiomic features analysis in computed tomography images of lung nodule classification. PLoS One 13:e0192002
21. Hatt M, Tixier F, Pierce L, Kinahan PE, Le Rest CC, Visvikis D (2016) Characterization of PET/CT images using texture analysis: the past, the present… any future? Eur J Nucl Med Mol Imaging 44: 151–165

22. Soh LK, Tsatsoulis C (1999) Texture analysis of SAR sea ice imagery using gray level co- occurrence matrices. IEEE Trans Geosci Remote Sens 37:780–795
23. Tibshirani R (1996) Regression shrinkage and selection via the lasso. J R Stat Soc Series B Stat Methodol 58:267–288
24. Friedman J, Hastie T, Tibshirani R (2010) Regularization paths for generalized linear models via coordinate descent. J Stat Softw 33:1–22
25. Liu Y, Wang H, Li Q et al (2018) Radiologic features of small pulmonary nodules and lung cancer risk in the National Lung Screening Trial: a nested case-control study. Radiology 286:298–306
26. Armato SG 3rd, McNitt-Gray MF, Reeves AP et al (2007) The Lung Image Database Consortium (LIDC): an evaluation of radiologist variability in the identification of lung nodules on CT scans. Acad Radiol 14:1409–1421
27. Kalpathy-Cramer J, Zhao B, Goldgof D et al (2016) A comparison of lung nodule segmentation algorithms: methods and results from a multi-institutional study. J Digit Imaging 29:476–487
28. Herder GJ, van Tinteren H, Golding RP et al (2005) Clinical prediction model to characterize pulmonary nodules: validation and added value of 18F-fluorodeoxyglucose positron emission tomography. Chest 128:2490–2496
29. He L, Huang Y, Ma Z, Liang C, Liang C, Liu Z (2016) Effects of contrast-enhancement, reconstruction slice thickness and convolution kernel on the diagnostic performance of radiomics signature in solitary pulmonary nodule. Sci Rep 6:34921
30. 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
31. Larue RT, Defraene G, De Ruysscher D, Lambin P, Van Elmpt W (2017) Quantitative radiomics studies for tissue characterization: a review of technology and methodological procedures. Br J Radiol 90:20160665
32. Zhao B, Tan Y, Tsai WY, Schwartz LH, Lu L (2014) Exploring variability in CT characterization of tumors: a preliminary phantom study. Transl Oncol 7:88–93
33. Zhao B, Tan Y, Tsai WY et al (2016) Reproducibility of radiomics for deciphering tumor phenotype with imaging. Sci Rep 6:23428
34. Lo P, Young S, Kim HJ, Brown MS, McNitt-Gray MF (2016) Variability in CT lung-nodule quantification: effects of dose reduction and reconstruction methods on density and texture based features. Med Phys 43:4854–4865
35. Lu L, Ehmke RC, Schwartz LH, Zhao B (2016) Assessing agreement between radiomic features computed for multiple CT imaging settings. PLoS One 11:e0166550
36. Shafiq-ul-Hassan M, Zhang GG, Latifi K et al (2017) Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. Med Phys 44:1050–1062
37. Mazzone PJ, Silvestri GA, Patel S et al (2018) Screening for lung cancer: CHEST guideline and expert panel report. Chest 153:954–985
38. Hastie T, Tibshirani R, Sherlock G, Eisen M, Brown P, Botstein D (1999) Imputing missing data for gene expression arrays. Stanford University Statistics Department Technical report. URL: http://www-stat.stanford.edu/~hastie/Papers/missing.pdf. Last downloaded 2019-03-01