



# Integrating manual diagnosis into radiomics for reducing the false positive rate of $^{18}\text{F}$ -FDG PET/CT diagnosis in patients with suspected lung cancer

Fei Kang<sup>1</sup> · Wei Mu<sup>2,3</sup> · Jie Gong<sup>4</sup> · Shengjun Wang<sup>1</sup> · Guoquan Li<sup>1</sup> · Guiyu Li<sup>1</sup> · Wei Qin<sup>4</sup> · Jie Tian<sup>2,4</sup> · Jing Wang<sup>1</sup> 

Received: 20 March 2019 / Accepted: 26 June 2019 / Published online: 18 July 2019  
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

## Abstract

**Purpose** The high false positive rate (FPR) of  $^{18}\text{F}$ -FDG PET/CT in lung cancer screening represents a severe challenge for clinical decision-making. This study aimed to develop a clinical-translatable radiomics nomogram for reducing the FPR of PET/CT in lung cancer diagnosis, and to determine the impact of integrating manual diagnosis to the performance of the radiomics nomogram.

**Methods** Among 3,947  $^{18}\text{F}$ -FDG PET/CT-screened patients with lung lesion, 157 malignant and 111 benign patients were retrospectively enrolled and divided into training and test cohorts. The data of manual diagnosis were recorded. A total of 4,338 features were extracted from CT, thin-section CT, PET and PET/CT, and the four radiomics signatures (RS) were then generated by LASSO method. Radiomics prediction nomogram integrating imaging-based RS and manual diagnosis was developed using multivariable logistic regression. The performances of RS and prediction nomograms were independently validated through key discrimination index and clinical benefit.

**Results** The FPR of manual diagnosis was found to be 30.6%. Among the four RS, PET/CT RS exhibited the best performance. By integrating manual diagnosis, the hybrid nomogram integrating PET/CT RS and manual diagnosis demonstrated lowest FPR and highest area under curve (AUC) and Youden index (YI) in both training and test cohorts (FPR: 5.4% and 9.1%, AUC: 0.98 and 0.92, YI: 85.8% and 75.5%, respectively). This hybrid nomogram respectively corrected 78.6% and 37.5% among FPR cases produced by PET/CT RS, without significantly sacrificing its sensitivity. The net benefit of hybrid nomogram appeared highest at <85% threshold probability.

**Conclusion** The established hybrid nomogram integrating PET/CT RS and manual diagnosis can significantly reduce FPR, improve diagnostic accuracy and enhance clinical benefit compared to manual diagnosis. By integrating manual diagnosis, the performance of this hybrid nomogram is superior to PET/CT RS, indicating the importance of clinicians' judgement as an essential information source for improving radiomics diagnostic approaches.

---

Fei Kang, Wei Mu and Jie Gong contributed equally to this work.

---

This article is part of the Topical Collection on Advanced Image Analyses (Radiomics and Artificial Intelligence).

---

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s00259-019-04418-0>) contains supplementary material, which is available to authorized users.

---

✉ Wei Qin  
wqin@xidian.edu.cn

✉ Jie Tian  
jie.tian@ia.ac.cn

✉ Jing Wang  
wangjing@fmmu.edu.cn

<sup>2</sup> Key Laboratory of Molecular Imaging of Chinese Academy of Sciences, Institute of Automation, Chinese Academy of Sciences, No. 95 Zhongguancun East Road, Beijing, China

<sup>3</sup> Department of Cancer Physiology, H. Lee Moffitt Cancer Center, Tampa, FL, USA

<sup>4</sup> Life Sciences Research Center, School of Life Sciences and Technology, Xidian University, No. 266 Xifeng Road, Xi'an, China

<sup>1</sup> Department of Nuclear Medicine, Xijing Hospital, Fourth Military Medical University, No.127 West Changle Road, Xi'an, China

**Keywords** Radiomics ·  $^{18}\text{F}$ -FDG PET/CT · False positive rate · Manual diagnosis · Lung lesion differentiation

## Introduction

Lung cancer is a leading cause of cancer deaths globally [1]. Medical imaging has been recognized as an indispensable tool for the early diagnosis of lung cancer [2]. Combining the structural information of computed tomography (CT) and the metabolic information of positron emission tomography (PET), a dual-modality  $^{18}\text{F}$ -FDG PET/CT imaging has been widely accepted as an advanced technique for lung cancer detection, with superior accuracy to both PET and CT [3–5].

Nonetheless,  $^{18}\text{F}$ -FDG PET/CT still faces a severe challenge in distinguishing benign lesions from malignant lung tumors. In a meta-analysis with large sample size, the false-positive rate (FPR) of  $^{18}\text{F}$ -FDG PET/CT for suspicious lung lesion is 25% on average, and as high as 39% in regions with endemic infectious lung disease [6]. In a multicenter study, the FPR in patients with high-risk lung nodules reaches up to 60.2% [7]. Additionally, due to the non-negligible error rate of biopsy histology (23%) [8] and the co-existence of infection and cancer [9], negative biopsy results are not sufficient for clinical physicians to confidently rule out the possibility of malignancy. The final confirmation of false-positive cases is limited by lack of definitive diagnostic approach. Only repeated biopsy and long-term follow-up [2], or imperfect diagnostic treatment [10] can be used to indirectly obtain a relatively accurate diagnosis. Such clinical dilemmas may result in additional costs of follow-up and potentially inaccurate clinical decisions, which remain large obstacles for PET/CT to play its due role in the early diagnosis of lung cancer.

The main reason for the high FPR of manual diagnosis can be attributed to two points. In the aspect of CT structural information, infectious lung lesions often mimic some key radiographic features of malignancy, including spicule sign and lobulation sign [11]. In the aspect of PET metabolic information, the activated immune cells in inflammatory site (e.g. tuberculosis) can result in an elevated level of glucose consumption, which also mimics the features of malignancy [12]. The above confusing macroscopic imaging manifestations may lead to a significant misguidance to the clinical experience-based differentiation of lung lesions [9]. Although many efforts have been made, such as delayed image acquisition and new PET tracers [13, 14], their actual clinical effects are still uncertain. Hence, developing new assistance methods for clinicians to reduce FPR is the key to solve the above dilemma.

As proposed in 2012, “radiomics” is a next-generation computer-aid diagnostic (CAD) technique for extracting a large number of image features from radiation images via a high-throughput approach and constructing nomograms based on machine-learning algorithms [15, 16]. Since medical

images are considered not merely as pictures for visual assessment but rather as mineable quantitative data, a radiomics approach can exert a great ability to improve diagnostic accuracy [8, 17–19]. Compared to previously used lung texture analysis, radiomics analysis generates more multidimensional information, such as intensity histogram, tumor shape, texture patterns, multiscale wavelets, tumor location and quantitative clinical biomarkers [16, 20], which serve an evolution of CAD with higher accuracy and wider application range [21]. Besides, radiological studies on CT have provided evidence that the discrimination capacity of the radiomics method on lung cancer is better than that of radiologists’ clinical experience [22, 23]. Thus, a radiomics nomogram may offer great potential to reduce FPR of  $^{18}\text{F}$ -FDG PET/CT in the differential diagnosis of lung lesions.

In recent years, PET/CT-based radiomics studies have emerged in the field of lung cancer, by focusing on algorithm methodology [24–26], survival prediction [27, 28] and tumor origin classification [29]. However, to our knowledge, study on the establishment and validation of PET/CT radiomics nomogram for reducing the FPR of lung cancer diagnosis has not yet been reported. Moreover, it remains unknown whether manual diagnosis based on clinical experience is a synergetic or misleading source of radiomics information. Therefore, in this study, we established four radiomics signatures (RS) of CT, thin-section CT, PET and PET/CT, and hybrid radiomics nomograms integrating RS and manual diagnosis. In addition, the performances of these RS and hybrid nomograms were compared, and the potential value of clinical experience integration was further determined.

## Materials and methods

This retrospective analysis was approved by the Ethics Committee of Xijing Hospital (Approval No. KY20173008–1), and the informed consent was waived. A total of 3,947  $^{18}\text{F}$ -FDG PET/CT-screened cases with lung lesion at Xijing Hospital from 2007 to 2017 were involved. Firstly, after excluding the patients (i) without obvious lung lesion, (ii) without documented pathological diagnosis, (iii) with a lesion diameter of <0.8 cm, (iv) with mainly non-solid component within the lesion (e.g. exudates, fibrosis and cavity), or (v) with chemotherapy and/or radiotherapy before PET/CT scans, 325 patients with suspicious solid lung lesion were preliminarily included for further pathological grouping. Secondly, according to the pathologic diagnosis and follow-up findings (diagnostic treatment, medical imaging scans or secondary biopsy), 111 patients were categorized into a benign group after excluding the potentially false-negative diagnosis caused

by sampling error. Thirdly, 157 patients with primary lung cancer were classified into malignant group based on biopsy and immunohistochemistry, by excluding the ones with pulmonary metastatic cancer from other origin. Finally, these patients were randomly divided into cohorts of training ( $n = 135$ ) and test ( $n = 133$ ). The workflow diagram of patient selection and grouping is shown in Supplementary Fig. S1.

## PET/CT and thin-section CT imaging

All the patients received  $^{18}\text{F}$ -FDG PET/CT and thin-section CT (TSCT) scans on the same scanner (Biograph 40, Siemens) by following a standard clinical protocol [30]. CT parameters were 100 kV, 110 mAs (CARE-Dose4D technology), 0.5-s rotation time, 5.0 mm ( $24 \times 1.2$  mm) slice thickness, 700 mm field of view, and  $512 \times 512$  matrix. PET parameters were 4.44–5.55 MBq/kg, 60 min after tracer administration, and 3 min per bed position. PET and CT images were reconstructed using an ordered-subsets expectation maximization algorithm with four iterations and eight subsets. TSCT parameters were 100 kV, 110 mAs, 0.5-s rotation time thickness, 700 mm field of view, and  $512 \times 512$  matrix.

## Manual diagnosis

In this retrospective study, manual diagnoses were obtained from the first cancerous or benign diagnosis of structured reports in the medical information system, taking no account of the secondary or subordinate diagnosis. Following the routine clinical workflow, the diagnoses of these structured reports were made through consensus by three experienced PET/CT clinicians, referring to the medical history, symptoms and other examination results. Furthermore, the sensitivity (SEN), specificity (SPE), positive predictive value (PPV), negative predictive value (NPV), false positive rate (FPR) and false negative rate (FNR) of manual diagnosis were calculated.

## Radiomics feature extraction

Three-dimensional regions of interest (3D-ROI) on the CT and PET images were respectively delineated by using ITK-SNAP software ([www.itksnap.org](http://www.itksnap.org)). All the delineation work was done under the guidance of experienced PET/CT clinicians blinded to the pathological grouping. CT 3D-ROI was manually delineated slice-by-slice around the lesions in CT images of lung window. For PET segmentation, 3D-ROI was semi-automatically drawn using the “adaptive brush” tool, by referencing to the 3D-ROI with a standard uptake value (SUV) threshold of 40% in the TrueD tool kit of Siemens MMWP workstation. For the lesions with anatomy

boundary close to mediastinum or bronchus, PET images were used as reference to determine the actual boundary in CT delineation.

A total of 4338 imaging radiomics features were respectively extracted from both PET and CT images using the MATLAB procedure algorithm, as described in the supplementary materials.

## Establishment of radiomics signatures (RS) and nomograms

After feature preselection according to the classification ability and redundancy (shown in supplementary materials in detail), we used a least absolute shrinkage and selection operator (LASSO) algorithm to select the most distinguishing features for each imaging modality (CT, TSCT, PET and PET/CT) as previously described [31, 32]. These selected features were weighted by their respective coefficients, and four different RS were then calculated using a linear combination of the features. Univariable logistical regression analysis was carried out on the four RS and the clinical factors in the training cohort. Only factors with statistically significant odds ratio (OR) were used to construct a prediction nomogram with multivariable logistic regression analysis for validation.

## Validation and comparison of RS and nomograms

Calibration curves were plotted to assess the linearity of each prediction nomogram, accompanied by the Akaike information criterion (AIC) value. A relatively corrected AUC value was calculated by bootstrapping validation (1000 bootstrap resamples). By using the cutoff value derived from receiver operating characteristic (ROC) curve analysis in the training cohort, the optimal diagnostic parameters [SEN, SPE, FPR, FNR, Youden’s index ( $YI = SEN + SPE - 1$ )] were calculated. The values of AUC and YI, which positively correlated with diagnostic performance, were used to quantitatively compare the predictive accuracy of the established nomograms. Decision curve analysis was implemented to determine the clinical benefit of RS and nomograms, following a previously reported protocol [33].

## Statistical analysis

All statistical analyses were performed using R software version 3.0.1 (<http://www.Rproject.org>). LASSO-logistic regression was analyzed by the “glmnet” package. The “RMS” package was used for multivariate binary logistic regression, nomograms and calibration. AUC calculation and decision

curve analysis was carried out with “Hmisc” and “dca.R” packages. To determine the agreement between the estimated probability and the actual malignant rate, calibration plots were obtained using the “calibrate” from the “rms” package. Decision curves were derived by using the “decision\_curve” from “rmda” package to quantify the net benefits at different threshold probabilities. Chi-squared test was used to compare the difference between categorical variables, while *t*-test was used for the comparison of two groups. A *P* value less than 0.05 was considered statistically significant.

## Results

### Clinical characteristics

As shown in Table 1, no significant differences were found between the training and test cohorts in terms of age, gender, pathological groups and manual diagnostic results ( $p > 0.05$ ). The balance between the two cohorts justified the rationality of patient grouping.

### Diagnostic performance of manual diagnosis

Analyzing the data shown in Table 1, the overall sensitivity, specificity, PPV and NPV of manual diagnosis were 93.6, 69.4,

81.2 and 88.5%, respectively. The overall FPR was found to be 30.6% (30.4% for training cohort and 30.9% for test cohort).

### Establishment of radiomics signatures (RS) and nomograms

Feature selection is shown in Supplementary Fig. S2. The formulas of RS are shown in Supplementary formulas I–IV. Detailed data on univariate and multivariate logistic regression analyses of all the RS and the prediction nomograms integrating RS and manual diagnosis are listed in Supplementary Tables S2–S4. Because of the best performance among the established four prediction nomograms, the nomogram that integrates PET/CT RS and manual diagnosis (hereinafter referred to as hybrid nomogram) was chosen for further analysis (Fig. 1a). Good calibration of this hybrid nomogram for predicting lung malignancy in training and test cohorts is respectively shown in Fig. 1b and c.

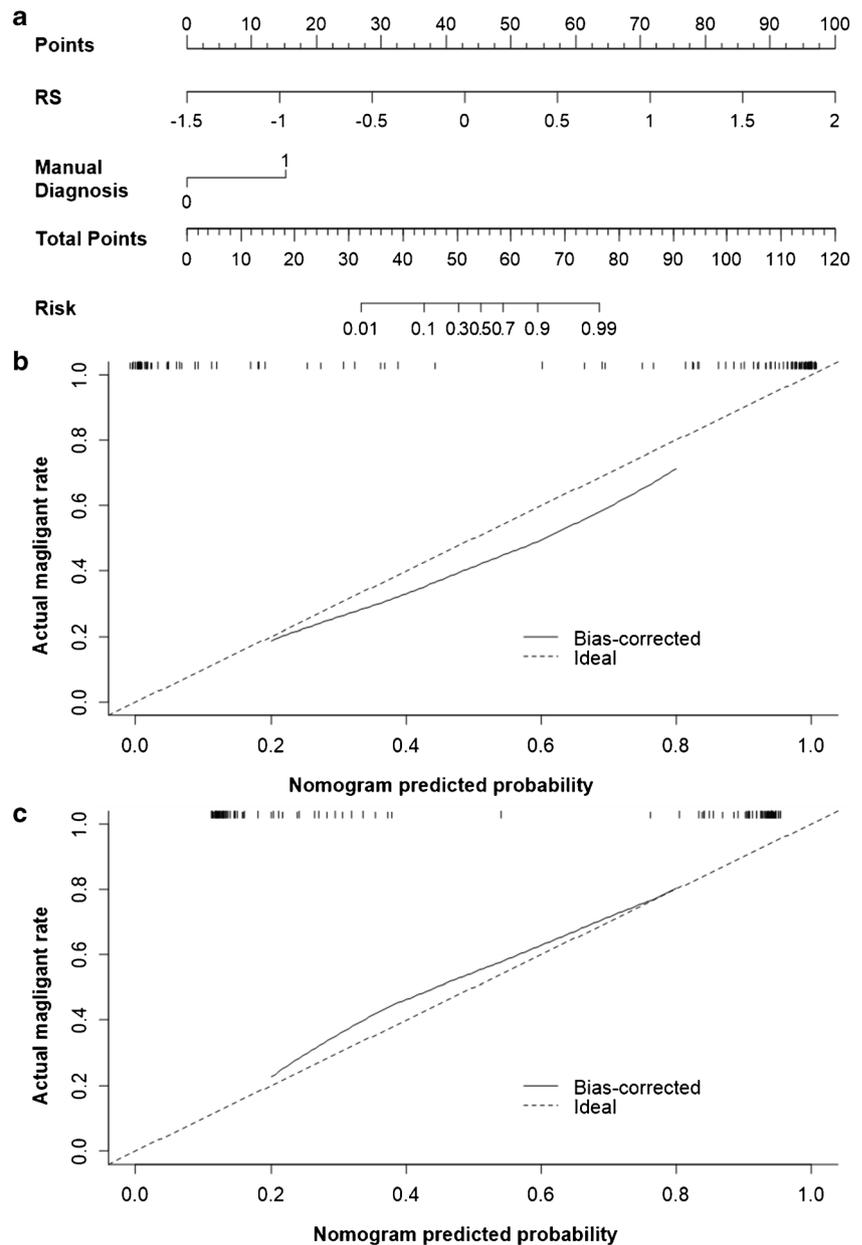
### Performance comparison of RS and nomograms

As shown in Fig. 2a and Supplementary Table S1, among the four imaging modalities, the RS' of malignant patients were significantly higher than those of benign patients in both training and test cohorts ( $p < 0.001$ ), indicating their potential ability for differentiating lung lesions. The ROC curves of the four RS and hybrid nomogram are presented in Fig. 2b. Among the

**Table 1** Characteristics of patients

Characteristic	Training cohort ( <i>n</i> = 135)		Test cohort ( <i>n</i> = 133)	
	Malignant ( <i>n</i> = 79)	Benign ( <i>n</i> = 56)	Malignant ( <i>n</i> = 78)	Benign ( <i>n</i> = 55)
Age, years (mean ± SD)	60.0 ± 12.6	51.3 ± 18.1	59.9 ± 10.2	51.4 ± 21.1
Gender				
Female	29 (36.7%)	29 (51.8%)	29 (37.2%)	28 (50.9%)
Male	50 (63.3%)	27 (48.2%)	49 (62.8%)	27 (49.1%)
Pathological category				
Squamous carcinoma	27 (34.2%)		25 (32.0%)	
Adenocarcinoma	41 (51.9%)		46 (59.0%)	
Small cell lung cancer	7 (8.9%)		6 (7.7%)	
Adenosquamous carcinoma	2 (2.5%)		0 (0.0%)	
Alveolar carcinoma	2 (2.5%)		1 (1.3%)	
Tuberculosis		20 (35.7%)		23 (41.8%)
Inflammation		35 (62.5%)		31 (56.4%)
Cryptococcal infection		1 (1.8%)		0
Systemic amyloidosis		0		1 (1.8%)
Manual diagnosis				
Malignant	75 (94.9%)	17 (30.4%)	72 (92.3%)	17 (30.9%)
Benign	4 (5.1%)	39 (69.6%)	6 (7.7%)	38 (69.1%)

**Fig. 1** (a) Establishment of the hybrid nomogram by integrating PET/CT radiomics signature (RS) and manual diagnosis. (b, c) Calibration curve of the hybrid nomogram in training cohort and test cohort. The *diagonal dotted line* represents a perfect estimated malignant rate by an ideal model. *Solid line* represents estimated malignant rate of the hybrid nomogram. Good alignment of these two lines indicates a good performance



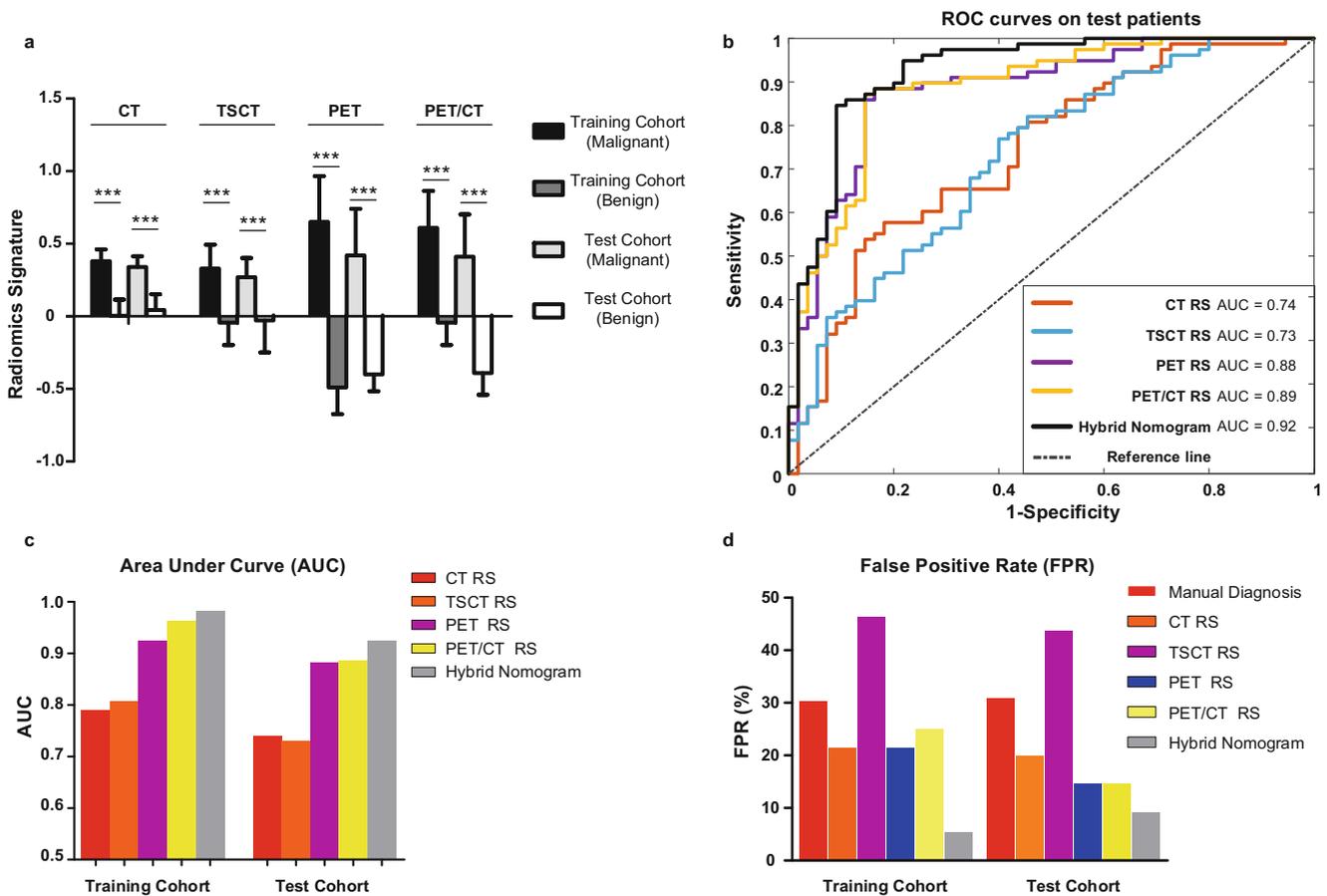
four RS, PET/CT RS showed the highest AUC. The AUC of hybrid nomogram was higher than that of PET/CT RS.

As shown in Table 2, the YI of hybrid nomogram was the highest, while FPR was the lowest in both training and test cohorts, indicating a high accuracy of hybrid nomogram in distinguishing lung lesions. Based on the cutoff values derived from ROC curves, the diagnostic thresholds of CT, TSCT, PET, PET/CT RS and hybrid nomogram were 0.6623, 0.4605, 0.4624, 0.3849 and 0.7179, respectively.

Compared to the manual diagnostic data shown in Table 2, the FPR of hybrid nomogram was remarkably decreased from 30.4 to 5.4% and 30.9 to 9.1% in the training cohort and test cohort, respectively, without generating new false positive

cases (a representative case corrected by hybrid nomogram is presented in Fig. 3a). Moreover, hybrid nomogram also, respectively, corrected one and two false-negative cases in training and test cohorts (a representative case is presented in Supplementary Fig. S3A). However, there were four and eight new false-negative cases, respectively, generated by hybrid nomogram in both training and test cohorts, leading to a moderate decrease in its sensitivity (94.9% vs. 91.1% for training cohort; 92.3% vs. 84.6% for test cohort).

Compared to PET/CT RS without integration of manual diagnosis, the hybrid nomogram exhibited higher AUC (training / test: 0.98 / 0.92 vs. 0.96 / 0.89), higher YI (training / test: 85.8% / 75.5% vs. 69.9% / 72.6%) and lower FPR (training / test: 5.4% / 9.1% vs. 25.0% / 14.6%), as shown in Fig. 2c. The



**Fig. 2** (a) Comparison of the RS of four imaging modalities in training and test cohorts. Data are presented as mean ± SD, \*\*\**P* < 0.001. (b) ROC curves of the four RS and hybrid nomogram in test cohort. (c, d) Comparison of AUC and FPR among the four RS and hybrid nomogram

hybrid nomogram respectively corrected 78.6% (11/14) and 37.5% (3/8) FPR among cases produced by PET/CT RS in training and test cohorts (a representative case is presented in Fig. 3b). However, the incorporation of manual diagnosis into the hybrid nomogram respectively generated four and two false-negative cases in both training and test cohorts, resulting in a slight reduction in its sensitivity (94.9% vs. 91.1% for training cohort; 87.2% vs. 84.6% for test cohort). A representative case is presented in Supplementary Fig. S3B. Moreover, by integrating manual diagnosis in the test cohort, the AUC of the CT-, TSCT- and PET-based prediction nomograms could also be respectively increased by 29.7%, 23.4% and 3.9% (data from Supplementary Tables S3 and S4).

**Decision curve analysis**

The decision curves for RS and hybrid nomogram are presented in Fig. 4. The net benefit of hybrid nomogram appeared highest when the threshold probability reached <85% (especially in the range of 60–85%), compared to the four RS.

**Discussion**

In this study, two major findings were highlighted. First, compared to manual diagnosis, the hybrid nomogram integrating PET/CT RS and manual diagnosis could reduce FPR from 30.6% to 5.4–9.1%, without affecting its sensitivity. Second, the hybrid nomogram achieved higher diagnostic accuracy, lower FPR and optimum clinical net benefit, compared to PET/CT RS without clinical judgment integration. These findings not only demonstrate the potential of radiomics technique in solving the clinical dilemma of high FPR in PET/CT-based lung cancer diagnosis, but also reveal the importance of clinical experience for improving radiomics performance.

Over the last two decades, researchers have made substantial efforts to apply prediction models in either a manual or computer-aided manner for lung lesion differentiation, mostly based on CT images [34–38]. A few studies have highlighted the great potential of PET-based texture analysis [39–41]. However, a limited number of texture features and lack of independent validation can influence the actual effectiveness and stability of CAD methods. In this study, we impartially selected features, including broader information of intensity,

**Table 2** Diagnostic performance of nomograms

Diagnostic models	Training cohort (95% CI)					Test cohort (95% CI)						
	AUC	SEN %	SPE %	FPR %	FNR %	YI %	AUC	SEN %	SPE %	FPR %	FNR %	YI %
CT RS	0.79 (0.71–0.87)	64.6 (53.8–75.3)	78.6 (67.7–89.5)	21.4 (10.5–32.3)	35.4 (24.4–46.2)	43.1 (21.5–64.8)	0.74 (0.66–0.83)	57.7 (46.6–68.8)	80.0 (69.6–90.4)	20.0 (9.6–30.4)	42.3 (31.2–53.4)	37.7 (16.2–59.2)
TSCT RS	0.81 (0.74–0.88)	89.9 (83.1–96.7)	53.6 (40.1–67.1)	46.4 (33.0–59.9)	10.1 (3.4–16.9)	43.4 (23.2–63.7)	0.73 (0.64–0.82)	79.5 (70.6–88.3)	56.4 (43.2–69.5)	43.6 (30.5–56.8)	20.5 (11.7–29.4)	35.9 (13.8–57.9)
PET RS	0.93 (0.89–0.96)	93.7 (88.3–99.1)	78.6 (68.0–89.2)	21.4 (10.8–32.0)	6.3 (9.2–11.7)	72.2 (56.2–88.2)	0.88 (0.82–0.94)	85.9 (78.0–93.8)	85.5 (76.1–94.9)	14.6 (5.2–23.9)	14.1 (6.2–22.0)	71.4 (54.0–88.7)
PET/CT RS	0.96 (0.94–0.99)	94.9 (90.0–99.9)	75.0 (63.5–86.5)	25.0 (13.5–36.5)	5.1 (0.9–10.0)	69.9 (53.5–86.4)	0.89 (0.83–0.94)	87.2 (79.5–94.9)	85.5 (76.5–94.5)	14.6 (5.5–23.6)	12.8 (5.1–20.6)	72.6 (55.9–89.4)
Hybrid nomogram	0.98 (0.97–1.00)	91.1 (84.9–97.4)	94.6 (88.7–100.0)	5.4 (0.0–11.3)	8.9 (2.6–15.2)	85.8 (73.5–98.0)	0.92 (0.88–0.97)	84.6 (76.6–92.6)	90.9 (83.3–98.6)	9.1 (1.4–16.8)	15.4 (7.4–23.4)	75.5 (59.9–91.2)
Manual diagnosis		94.9	69.6	30.4	5.1	64.6		92.3	69.1	30.9	7.7	61.4

**Note:** RS: radiomics signature; SEN: sensitivity; SPE: specificity; FPR: false positive rate; FNR: false negative rate; YI: Youden index; CI: confidence interval

shape, texture and wavelet, from 4338 multi-dimensional features, following a radiomics protocol. In addition, the performances of RS and hybrid nomogram were evaluated by independent validations.

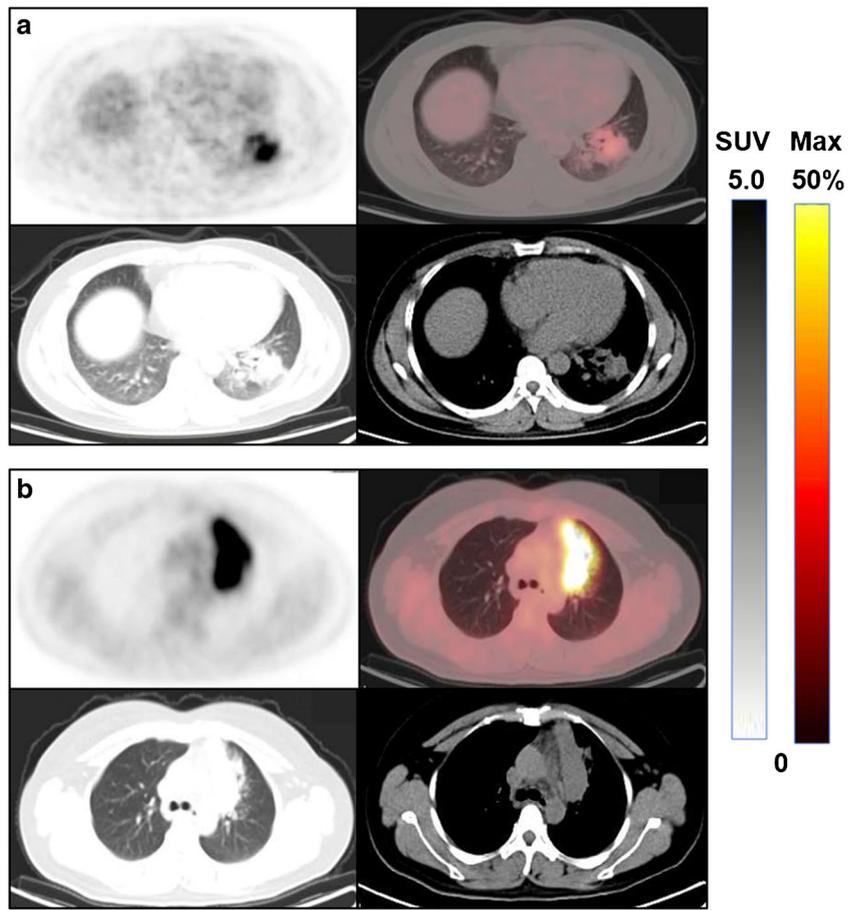
In addition, our study provides evidence that clinical experience may exert remarkably synergetic effect on PET/CT RS with regard to FPR reduction and diagnostic accuracy enhancement. In previous studies, clinical diagnosis and CAD methods are often set as competitors in the race of predictive accuracy. Nonetheless, clinical experience-based manual diagnosis represents an integration of clinical and imaging information, in a brain-based intelligent way. Our results imply that clinical experience is not likely to be obsolete or dispensable in the presence of artificial intelligence diagnosis, but remains a crucial player for establishing better diagnostic method.

After analyzing the established RS formulas, a looming link between the major differentiating PET features and clinical experience appears to exist. The major differentiating PET features derived from the original plane in the formulas could partly agree with the general principles of clinical experience, in which the malignant possibility is positively correlated with non-uniformity and PET signal value. For instance, the PET features *W1.VAR\_11* and *W1.LRHGLE\_13*, which respectively are associated with the degree of variation and the proportion of high value-pixels, are positively correlated to malignancy. Meanwhile, the PET feature *W1.RP\_2*, which related to the uniformity degree of equal value-pixels, is negatively correlated to malignancy. At present, one of the major barriers to the CAD method application is the limited theoretical interpretability of these prediction models [19, 42]. Our results suggest that the radiomics models may confer an underlying theoretical consistency with clinical experience. Although this claim is too preliminary and the detailed intuitive meaning of radiomics features (especially wavelet feature) remains enigmatic, more studies are warranted for establishing a theoretical connection between radiomics mechanism and clinical experience, to provide a deeper understanding on radiomics and to discover new radiological principles.

In the aspect of clinical benefit, the hybrid nomogram exhibits highest net benefit when the threshold probability was <85%, especially in the range of 60–85%. According to the principle of decision curve, the threshold-probability value largely depends on the prevalence of positive results in a population [43]. Previous meta-analyses have shown that the median prevalence of malignancy in lung nodule PET/CT cases ranged from 60 to 72.5% [6, 44], fitting the preferred range of hybrid nomogram. Furthermore, our results demonstrated the practical clinical value of this hybrid nomogram for differentiating lung cancer patients.

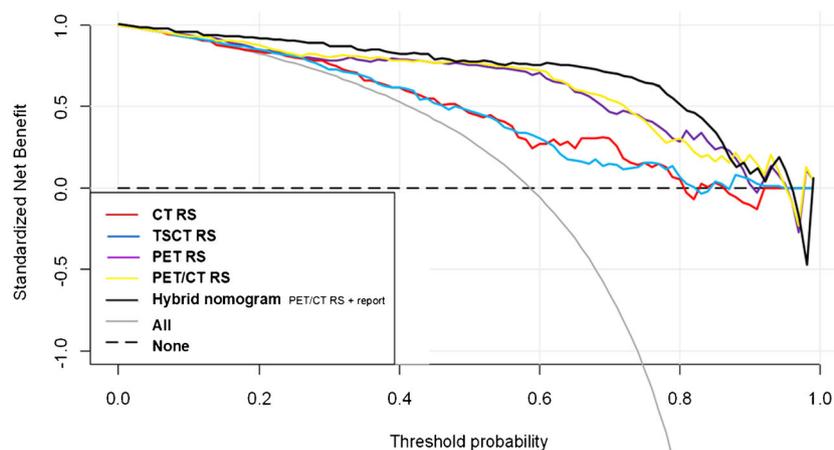
This study has several limitations. Firstly, this retrospective study investigated patients from a single center, and thus multicenter prospective studies are warranted. Considering that

**Fig. 3** Transaxial <sup>18</sup>F-FDG PET, PET/CT, CT images and basic characteristics of representative cases. **(a)** A representative false-positive case (#b106) corrected by hybrid nomogram. Male, postoperative pathology confirmed tuberculosis in the left lung after pulmonary segmentectomy. The manual diagnosis was malignant for positive SUVmax/mean (5.9/2.8), spicule sign and pleural stretch sign. The result of hybrid nomogram was 0.0737 for this case (malignancy threshold: 0.7179). **(b)** A representative case (#b83) showing the impact of clinical experience integration. A female patient with tuberculosis in the left lung confirmed by puncture biopsy and follow-up. The manual diagnosis was tuberculosis for the location, symptom and regular shape of the lesion. The value of PET/CT RS was 0.8769 (malignancy threshold: 0.3849). However, the value of hybrid nomogram was 0.4390 (malignancy threshold: 0.7179), indicating a benign lesion



the radiomics features are very sensitive to acquisition and reconstruction procedures [45, 46], a certain degree of standardization is a necessary precondition for multicenter studies. Secondly, this hybrid nomogram was solely based on RS and

manual diagnosis. Therefore, additional clinical information, such as serum biomarkers, may further enhance the performance of radiomics nomogram. Thirdly, the performances of CT- and TSCT-based RS are relatively poor in this study,



**Fig. 4** Decision curves of the four RS and hybrid nomogram. *Black line* represents hybrid nomogram incorporating PET/CT RS and manual diagnosis. *Red, blue, fuchsia and yellow lines* represent the RS of CT, TSCT, PET and PET/CT, respectively. *Grey reference line* indicates the assumption that all patients are diagnosed to be malignant, whereas the *black dashed reference line* indicates the assumption that all patients are

diagnosed to be benign. The hybrid nomogram exhibits highest net benefit when the threshold probability is <85%, especially in the range of 60–85%. This range fits the reported median prevalence of malignancy in PET/CT lung nodule cases (60–72.5% [6, 44]), indicating the practical clinical value of the hybrid nomogram

probably due to the sub-optimal CT and TSCT imaging quality of the standard PET/CT imaging protocol.

**Acknowledgments** We would like to thank Zhe Wang, Zhiyong Quan, Ni Wang, Jingwei Yi, Qingju Zhang, Jin Zeng and Xiaohu Zhao for their technical assistance. Fei Kang thanks Prof. Yaochi Yang and Shuangqin Wu for providing the necessary impetus to conduct this study.

**Funding** This work was supported by the National Natural Science Foundation of China (Grant Nos. 81871379, 816771713, 81601521), the National Key R&D Program of China (Grant No. 2016YFC0103804), and the Young Elite Scientists Sponsorship Program of China Association for Science and Technology (Grant No. 2017QNRC001).

## Compliance with ethical standards

**Conflict of interest** No other potential conflict of interest relevant to this article was reported.

**Ethical approval** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

**Informed consent** This retrospective analysis was approved by the Ethics Committee of Xijing Hospital (Approval No. KY20173008–1), and the informed consent was waived.

## References

- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68:394–424.
- Wood DE, Kazerooni EA, Baum SL, Eapen GA, Ettinger DS, Hou L, et al. Lung Cancer screening, version 3.2018, NCCN clinical practice guidelines in oncology. *J Natl Compr Cancer Netw*. 2018;16:412–41.
- Kim SK, Allen-Auerbach M, Goldin J, Fueger BJ, Dahlbom M, Brown M, et al. Accuracy of PET/CT in characterization of solitary pulmonary lesions. *J Nucl Med*. 2007;48:214–20.
- Kagna O, Solomonov A, Keidar Z, Bar-Shalom R, Fruchter O, Yigla M, et al. The value of FDG-PET/CT in assessing single pulmonary nodules in patients at high risk of lung cancer. *Eur J Nucl Med Mol Imaging*. 2009;36:997–1004.
- Jeong SY, Lee KS, Shin KM, Bae YA, Kim BT, Choe BK, et al. Efficacy of PET/CT in the characterization of solid or partly solid solitary pulmonary nodules. *Lung Cancer*. 2008;61:186–94.
- Deppen SA, Blume JD, Kensinger CD, Morgan AM, Aldrich MC, Massion PP, et al. Accuracy of FDG-PET to diagnose lung cancer in areas with infectious lung disease: a meta-analysis. *JAMA*. 2014;312:1227–36.
- Maiga AW, Deppen SA, Mercaldo SF, Blume JD, Montgomery C, Vaszar LT, et al. Assessment of Fluorodeoxyglucose F18-labeled positron emission tomography for diagnosis of high-risk lung nodules. *JAMA Surg*. 2018;153:329–34.
- Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. 2016;278:563–77.
- Shetty N, Noronha V, Joshi A, Rangarajan V, Purandare N, Mohapatra PR, et al. Diagnostic and treatment dilemma of dual pathology of lung cancer and disseminated tuberculosis. *J Clin Oncol*. 2014;32:e7–9.
- Glasziou P, Rose P, Heneghan C, Balla J. Diagnosis using "test of treatment". *BMJ*. 2009;338:b1312.
- Li Y, Su M, Li F, Kuang A, Tian R. The value of (1)(8)F-FDG-PET/CT in the differential diagnosis of solitary pulmonary nodules in areas with a high incidence of tuberculosis. *Ann Nucl Med*. 2011;25:804–11.
- Haroon A, Zumla A, Bomanji J. Role of fluorine 18 fluorodeoxyglucose positron emission tomography-computed tomography in focal and generalized infectious and inflammatory disorders. *Clin Infect Dis*. 2012;54:1333–41.
- Kim IJ, Lee JS, Kim SJ, Kim YK, Jeong YJ, Jun S, et al. Double-phase 18F-FDG PET-CT for determination of pulmonary tuberculoma activity. *Eur J Nucl Med Mol Imaging*. 2008;35:808–14.
- Kang F, Wang S, Tian F, Zhao M, Zhang M, Wang Z, et al. Comparing the diagnostic potential of 68Ga-Alfatide II and 18F-FDG in differentiating between non-small cell lung cancer and tuberculosis. *J Nucl Med*. 2016;57:672–7.
- Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. 2017;14:749–62.
- Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. 2012;48:441–6.
- Hatt M, Tixier F, Visvikis D, Cheze Le Rest C. Radiomics in PET/CT: more than meets the eye? *J Nucl Med*. 2017;58:365–6.
- Aerts H. Data science in radiology: a path forward. *Clin Cancer Res*. 2018;24:532–4.
- Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, et al. Artificial intelligence in cancer imaging: clinical challenges and applications. *CA Cancer J Clin*. 2019;69:127–57.
- Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*. 2014;5:4006.
- Hochhegger B, Zanon M, Altmayer S, Pacini GS, Balbinot F, Francisco MZ, et al. Advances in imaging and automated quantification of malignant pulmonary diseases: a state-of-the-art review. *Lung*. 2018;196:633–42.
- Perandini S, Soardi GA, Motton M, Augelli R, Dallaserra C, Puntel G, et al. Enhanced characterization of solid solitary pulmonary nodules with Bayesian analysis-based computer-aided diagnosis. *World J Radiol*. 2016;8:729–34.
- Ma JC, Wang Q, Ren YC, Hu HB, Zhao J. Automatic lung nodule classification with radiomics approach. *SPIE*; 2016; <https://doi.org/10.1117/12.2220768>.
- Desseroit MC, Tixier F, Weber WA, Siegel BA, Cheze Le Rest C, Visvikis D, et al. Reliability of PET/CT shape and heterogeneity features in functional and morphologic components of non-small cell lung Cancer tumors: a repeatability analysis in a prospective multicenter cohort. *J Nucl Med*. 2017;58:406–11.
- Sollini M, Cozzi L, Antunovic L, Chiti A, Kirienko M. PET Radiomics in NSCLC: state of the art and a proposal for harmonization of methodology. *Sci Rep*. 2017;7:358.
- Hatt M, Laurent B, Fayad H, Jaouen V, Visvikis D, Le Rest CC. Tumour functional sphericity from PET images: prognostic value in NSCLC and impact of delineation method. *Eur J Nucl Med Mol Imaging*. 2018;45:630–41.
- Arshad MA, Thornton A, Lu H, Tam H, Wallitt K, Rodgers N, et al. Discovery of pre-therapy 2-deoxy-2-(18)F-fluoro-D-glucose positron emission tomography-based radiomics classifiers of survival outcome in non-small-cell lung cancer patients. *Eur J Nucl Med Mol Imaging*. 2019;46:455–66.

28. Kirienko M, Cozzi L, Antunovic L, Lozza L, Fogliata A, Voulaz E, et al. Prediction of disease-free survival by the PET/CT radiomic signature in non-small cell lung cancer patients undergoing surgery. *Eur J Nucl Med Mol Imaging*. 2018;45:207–17.
29. Kirienko M, Cozzi L, Rossi A, Voulaz E, Antunovic L, Fogliata A, et al. Ability of FDG PET and CT radiomics features to differentiate between primary and metastatic lung lesions. *Eur J Nucl Med Mol Imaging*. 2018;45:1649–60.
30. Delbeke D, Coleman RE, Guiberteau MJ, Brown ML, Royal HD, Siegel BA, et al. Procedure guideline for tumor imaging with 18F-FDG PET/CT 1.0. *J Nucl Med*. 2006;47:885–95.
31. Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. *J Clin Oncol*. 2016;34:2157–64.
32. Liu Z, Zhang XY, Shi YJ, Wang L, Zhu HT, Tang Z, et al. Radiomics analysis for evaluation of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *Clin Cancer Res*. 2017;23:7253–62.
33. Vickers AJ, Cronin AM, Elkin EB, Gonen M. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. *BMC Med Inform Decis Mak*. 2008;8:53.
34. Rubin GD, Lyo JK, Paik DS, Sherbondy AJ, Chow LC, Leung AN, et al. Pulmonary nodules on multi-detector row CT scans: performance comparison of radiologists and computer-aided detection. *Radiology*. 2005;234:274–83.
35. Gould MK, Ananth L, Barnett PG, Veterans Affairs SCSG. A clinical model to estimate the pretest probability of lung cancer in patients with solitary pulmonary nodules. *Chest*. 2007;131:383–8.
36. Soardi GA, Perandini S, Motton M, Montemezzi S. Assessing probability of malignancy in solid solitary pulmonary nodules with a new Bayesian calculator: improving diagnostic accuracy by means of expanded and updated features. *Eur Radiol*. 2015;25:155–62.
37. Digumarthy SR, Padole AM, Lo Gullo R, Singh R, Shepard JO, Kalra MK. CT texture analysis of histologically proven benign and malignant lung lesions. *Medicine (Baltimore)*. 2018;97:e11172.
38. Choi W, Oh JH, Riyahi S, Liu CJ, Jiang F, Chen W, et al. Radiomics analysis of pulmonary nodules in low-dose CT for early detection of lung cancer. *Med Phys*. 2018;45:1537–49.
39. Nie Y, Li Q, Li F, Pu Y, Appelbaum D, Doi K. Integrating PET and CT information to improve diagnostic accuracy for lung nodules: a semiautomatic computer-aided method. *J Nucl Med*. 2006;47:1075–80.
40. Chen S, Harmon S, Perk T, Li X, Chen M, Li Y, et al. Diagnostic classification of solitary pulmonary nodules using dual time (18)F-FDG PET/CT image texture features in granuloma-endemic regions. *Sci Rep*. 2017;7:9370.
41. Miwa K, Inubushi M, Wagatsuma K, Nagao M, Murata T, Koyama M, et al. FDG uptake heterogeneity evaluated by fractal analysis improves the differential diagnosis of pulmonary nodules. *Eur J Radiol*. 2014;83:715–9.
42. Chen JH, Asch SM. Machine learning and prediction in medicine - beyond the peak of inflated expectations. *N Engl J Med*. 2017;376:2507–9.
43. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ*. 2016;352:i6.
44. Gould MK, Maclean CC, Kuschner WG, Rydzak CE, Owens DK. Accuracy of positron emission tomography for diagnosis of pulmonary nodules and mass lesions: a meta-analysis. *JAMA*. 2001;285:914–24.
45. Kim H, Park CM, Gwak J, Hwang EJ, Lee SY, Jung J, et al. Effect of CT Reconstruction Algorithm on the Diagnostic Performance of Radiomics Models: A Task-Based Approach for Pulmonary Subsolid Nodules. *AJR Am J Roentgenol*. 2019;212:505–12.
46. Vallieres M, Laberge S, Diamant A, El Naqa I. Enhancement of multimodality texture-based prediction models via optimization of PET and MR image acquisition protocols: a proof of concept. *Phys Med Biol*. 2017;62:8536–65.

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