

ORIGINAL ARTICLE

Development of a deep learning-based method to diagnose pulmonary ground-glass nodules by sequential computed tomography imaging

Zhixin Qiu^{1†}  | Qingxia Wu^{2†} | Shuo Wang^{3,4†} | Zhixia Chen⁵ | Feng Lin⁶ | Yuyan Zhou¹ | Jing Jin¹ | Jinghong Xian⁷ | Jie Tian^{2,3,4} | Weimin Li¹

¹Department of Respiratory and Critical Care Medicine, West China Hospital, Sichuan University, Chengdu, China

²College of Medicine and Biomedical Information Engineering, Northeastern University, Shenyang, China

³CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Beijing, China

⁴Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, School of Medicine and Engineering, Beihang University, Beijing, China

⁵Department of Radiology, West China Hospital, Sichuan University, Chengdu, China

⁶Department of Thoracic Surgery, West China Hospital, Sichuan University, Chengdu, China

⁷Department of Clinical Research, West China Hospital, Sichuan University, Chengdu, China

Correspondence

Jie Tian, CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, No.95 Zhongguancun east road, Beijing 100190, China.
Email: jie.tian@ia.ac.cn

Weimin Li, Department of Respiratory and Critical Care Medicine, West China Hospital, Sichuan University, Chengdu, 610041, China.
Email: weimi003@scu.edu.cn

Funding information

Central Government Guided Local Science and Technology Free Exploration Projects of Sichuan Province, China, Grant/Award Number: 2020ZYD005; China Postdoctoral Science Foundation, Grant/Award Number: 2019TQ0019; National Key R&D Program of China, Grant/Award Number: 2017YFA0205200; National Natural Science Foundation of China, Grant/Award Numbers: 81700095, 91859203, 81871890, 81930053; Science and Technology Plan Project of Sichuan Province, China, Grant/Award Number: 2019YFS0335

Abstract

Background: Early identification of the malignant propensity of pulmonary ground-glass nodules (GGNs) can relieve the pressure from tracking lesions and personalized treatment adaptation. The purpose of this study was to develop a deep learning-based method using sequential computed tomography (CT) imaging for diagnosing pulmonary GGNs.

Methods: This diagnostic study retrospectively enrolled 762 patients with GGNs from West China Hospital of Sichuan University between July 2009 and March 2019. All patients underwent surgical resection and at least two consecutive time-point CT scans. We developed a deep learning-based method to identify GGNs using sequential CT imaging on a training set consisting of 1524 CT sections from 508 patients and then evaluated 256 patients in the testing set. Afterwards, an observer study was conducted to compare the diagnostic performance between the deep learning model and two trained radiologists in the testing set. We further performed stratified analysis to further relieve the impact of histological types, nodule size, time interval between two CTs, and the component of GGNs. Receiver operating characteristic (ROC) analysis was used to assess the performance of all models.

Results: The deep learning model that used integrated DL-features from initial and follow-up CT images yielded the best diagnostic performance, with an area under the curve of 0.841. The observer study showed that the accuracies for the deep learning model, junior radiologist, and senior radiologist were 77.17%, 66.89%, and 77.03%, respectively. Stratified analyses showed that the deep learning model and radiologists exhibited higher performance in the subgroup of nodule sizes larger than 10 mm. With a longer time interval between two CTs, the deep learning model yielded higher diagnostic accuracy, but no general rules were yielded for

[†]Equal contribution.

radiologists. Different densities of components did not affect the performance of the deep learning model. In contrast, the radiologists were affected by the nodule component.

Conclusions: Deep learning can achieve diagnostic performance on par with or better than radiologists in identifying pulmonary GGNs.

KEYWORDS

deep learning, ground-glass nodules, multiple timepoints, sequential

INTRODUCTION

Ground-glass nodules (GGNs) are a nonspecific finding on chest CT that may occur in a variety of pulmonary diseases, such as malignancy, atypical adenomatous hyperplasia, inflammatory reaction, granuloma, and fibrosis, and are pathologically characterized by thickening of the alveolar wall and alveoli almost filled with exudate and a few lymphocytes, neutrophils, macrophages or tumor cells.^{1,2} The advent of low-dose spiral computer tomography (LDCT) has led to advancements in early screening for lung cancer, especially early-stage (stage IA) lung cancers presenting as GGNs.^{3–7} The detection rate of lung nodules in patients at high risk for lung cancer is approximately 27.9%,⁸ and the majority of cases diagnosed as early-stage lung cancer are subcentimeter GGNs.⁹ Henschke et al. analyzed 233 patients with lung nodules detected on LDCT and found that the malignancy rates for GGNs and solid nodules were 34.1% and 7%, respectively, indicating that the malignancy rate in GGNs is significantly higher than that in solid nodules.¹⁰ Although Kobayashil et al. proposed that GGNs should be followed for at least 3 years, they may become malignant during long-term follow-up, allowing the best time for intervention slip.¹¹ Moreover, repeated examinations will cause huge economic and psychological burdens for patients.¹² However, limited by existing imaging and other techniques, it is still very difficult to distinguish benign from malignant GGNs. Therefore, early accurate diagnosis and intervention of malignant GGNs will benefit clinical practice.

Deep learning (DL), as an advanced artificial intelligence algorithm, can mine shallow image intensity and shape features as well as high-dimensional abstract information due to its layered structure.¹³ Recently, with the widespread use of convolutional neural networks,¹⁴ DL has shown expert-level analytic performance in the CT imaging analysis of lung diseases, such as cancer screening,^{15,16} lesion segmentation,¹⁷ and prediction of EGFR gene mutations.¹⁸ DL does not need a precise lesion contour, which is quite suitable for analyzing GGNs, of which the lesion boundary is blurred. In this study, we aimed to apply DL technology to sequential CT images to mine predictive information and identify the benign and malignant properties of GGNs and compare the DL method with trained radiologists. We present the following article in accordance with the STARD reporting checklist.

METHODS

Patients

The workflow of this study is shown in Figure 1. From July 2009 to March 2019, 762 consecutive patients with GGNs on chest CT who underwent surgical resection at West China Hospital of Sichuan University were enrolled in this research. All patients met the following inclusion criteria: (i) chest CT manifested as GGNs in the lungs, (ii) underwent surgical resection at our hospital and was confirmed by pathology, (iii) no previous history of treatment with radiotherapy or chemotherapy, and (iv) performed at least two consecutive time-point CT scans. We excluded patients if: (i) they had atelectasis, pneumonia, hilar enlargement or pleural effusion and other imaging findings, (ii) they had other severe diseases (severe cardiovascular, cerebrovascular or lung diseases), (iii) incomplete clinical data or inability to contact patients, and (iv) inability to obtain paraffin specimens, insufficient paraffin specimens, or inconsistent tissue type after reslicing HE sections. We randomly divided the dataset into training and testing sets at a ratio of 2:1. The recruitment pathway is shown in Fig. S1 in the Supplement.

Baseline information, including age, sex, smoking, extrapulmonary cancer history, family cancer history, nodule size, type, location, time interval between two CTs, and pathology, is shown in Table 1. Note that nodule size was measured by averaging the long- and short-axes on transverse sections of initial CT using lung window settings following the Fleischner Society guidelines (senior radiologist ZC with 10+ years of experience).¹⁹

Image acquisition and preprocessing

All patients underwent unenhanced chest CT scans, and all image data were reconstructed with a thickness of 1.0 to 1.5 mm (more scanning parameters are provided in eMethods 1 in the Supplement). To extract nodule information for analysis, one radiologist (Z.C.) used rectangular bounding boxes to contour the whole nodule on the largest cross-section. Based on this marked ROI, we expanded two images forwards and backwards. Then, we connected every three adjacent images to form the three-channel model to fit the following DL model. All ROIs were scaled to $64 \times 64 \times 3$ voxel size, and the voxel intensities were normalized to [0, 1]. In this study, when a

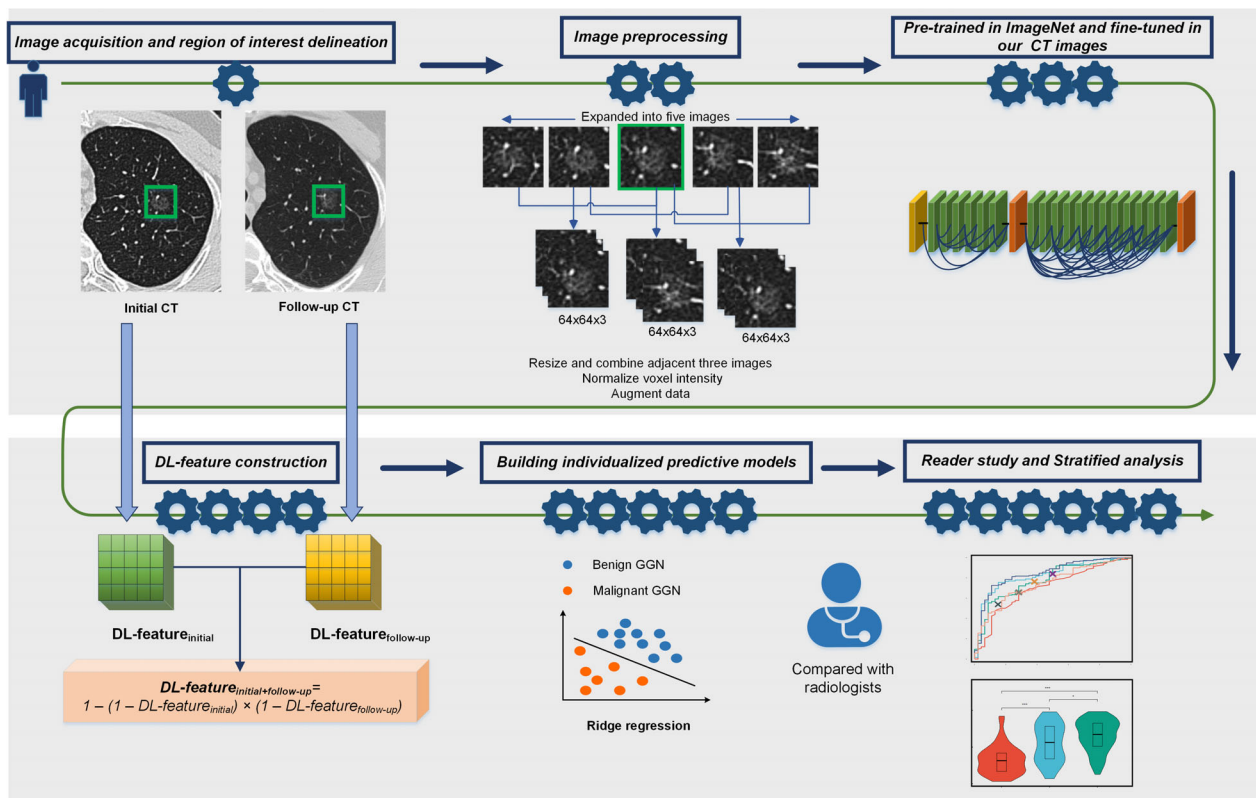


FIGURE 1 The workflow of this study. This study included the following six parts: (a) baseline CT and follow-up CT acquisition and ROI (the green box) delineation, (b) image preprocessing, (c) building a DL model that was pretrained in ImageNet and fine-tuned with our CT images, (d) constructing DL-features by initial CT and follow-up CT, (e) building individualized GGN prediction models by the DL-feature, and (f) comparing the DL model with radiologists

patient had multiple GGNs, only the one with the largest volume was enrolled.

Deep learning feature extraction

We developed a DL network to extract the intrinsic characteristics of GGNs from CT images at two time points. The network shared the same architecture with the first two dense blocks in DenseNet121.²⁰ The dense block represents the special structure in DenseNet that contributed to better performance than other networks. Each dense block was a stack of multiple convolutional, batch normalization layers with a rectified linear activation function, and the layers were directly connected to other layers in a feed-forward fashion (details of layers in eMethods 2 in the Supplement). This structure means that the network could combine the information between different convolutional layers, further benefiting optimization of the learning process. To strengthen the training process, we used the ImageNet dataset,²¹ which included 14 million natural images to pretrain the model and 1524 CT sections to fine-tune the model from the training set. When the loss the model on the training set was converged, we applied the weight on the testing set. To enlarge the training data and avoid

overfitting, we used data augmentation techniques on the fly during the training process, including random shift, translation, rotation, flipping, and zooming. We extracted the output of the last convolutional layer in the network and defined it as a DL-feature, which was 256-dimensional. Since every nodule included three $64 \times 64 \times 3$ input images, we averaged features to acquire the DL-feature for the nodule. The network was implemented and trained in Python 3.6 and Keras 2.2 (TensorFlow 1.7 backend).

We used the same network to extract the DL-feature from the initial and follow-up CT images. We defined the DL-feature extracted from the initial CT images as $DL\text{-feature}_{\text{initial}}$ and the DL-feature extracted from the follow-up CT images as $DL\text{-feature}_{\text{follow-up}}$.

Development of individualized predictive models

After gaining the DL-feature, we used ridge regression to build the individualized predictive model to distinguish benign from malignant GGNs. The ridge regression used L2 regularization to avoid overfitting and maintained as many original features as possible with a penalty parameter C . To gain the optimal C , we used 10-fold cross-validation in the

TABLE 1 Patient characteristics in the primary and validation cohorts

Characteristics	Training set (<i>n</i> = 508)		<i>p</i> ^a	Testing set (<i>n</i> = 254)		<i>p</i> ^a	<i>p</i> ^b 0.35
	Benign 108 (21.0)	Malignant 400 (79.0)		Benign 46 (22.7)	Malignant 208 (77.3)		
Age, years			< 0.001			0.007	0.08
Mean ± SD	49.6 ± 12.4	54.4 ± 11.2		50.7 ± 11.4	55.8 ± 10.9		
Sex (No. %)			0.34			0.75	0.67
Male	34 (31.5)	105 (26.2)		12 (26.1)	62 (29.8)		
Female	74 (68.5)	295 (73.8)		34 (73.9)	146 (70.2)		
Smoking (No. %)			0.77			0.22	0.25
Yes	12 (11.1)	51 (12.8)		4 (8.7)	36 (17.3)		
No	96 (88.9)	349 (87.2)		42 (91.3)	172 (82.7)		
Extrapulmonary cancer history (No. %)			0.006			0.37	0.37
Yes	2 (1.9)	44 (11.0)		3 (6.5)	26 (12.5)		
No	106 (98.1)	356 (89.0)		43 (93.5)	182 (87.5)		
Family cancer history (No. %)			0.17			0.06	0.30
Yes	10 (9.3)	60 (15.0)		3 (6.5)	40 (19.2)		
No	98 (90.7)	340 (85.0)		43 (93.5)	168 (80.8)		
Nodule size (No. %)			< 0.001			< 0.001	0.78
≤ 10 mm	73 (67.6)	180 (45.0)		36 (78.3)	94 (45.2)		
> 10 mm	35 (32.4)	220 (55.0)		10 (21.7)	114 (54.8)		
Type (No. %)			0.13			0.85	0.25
pGGN	59 (54.6)	183 (45.8)		23 (50.0)	110 (52.9)		
mGGN	49 (45.4)	217 (54.2)		23 (50.0)	98 (47.1)		
Location (No. %)			0.38			0.46	0.61
RUL	45 (41.7)	152 (38)		16 (8.7)	93 (44.7)		
RML	11 (10.2)	24 (6)		5 (10.9)	13 (6.2)		
RLL	18 (16.7)	63 (15.8)		5 (10.9)	25 (12.0)		
LUL	24 (22.2)	115 (28.7)		16 (34.8)	53 (25.5)		
LLL	10 (9.3)	46 (11.5)		4 (8.7)	24 (11.5)		
Time interval (No. %)			< 0.001			< 0.001	0.89
≤ 90 days	46 (42.6)	305 (76.2)		11 (23.9)	163 (78.4)		
(90, 180) days	25 (23.1)	68 (17.0)		19 (41.3)	26 (12.5)		
>180 days	37 (34.3)	27 (6.8)		16 (34.8)	19 (9.1)		
Pathology							
Inflammatory	73			37			
Granuloma	3			0			
Fibrosis	10			5			
Interstitial hyperplasia	19			4			
Hamartoma	1			0			
Sclerosing	1			0			
Tuberculosis	1			0			
Invasive adenocarcinoma ^c		325			180		
Preinvasive adenocarcinoma ^d		73			28		
Squamous carcinoma		2			0		

Note: Cancer (IASLC)⁴³

Abbreviations: LLL, left lower lobe; LUL, left upper lobe; mGGN, mixed ground-glass nodule; pGGN, pure ground-glass nodule; RLL, right lower lobe; RML, right middle lobe; RUL, right upper lobe.

^a*p* is derived from the univariable association analyses of each clinicopathological variable. Between patients with benign and malignant GGN in the training and testing set, respectively.

^b*p* represents the difference of each clinicopathological variable between the training and testing set.

^cInvasive adenocarcinoma includes minimally invasive adenocarcinoma and invasive pulmonary adenocarcinoma according to The International Association for the Study of Lung.

^dPreinvasive adenocarcinoma includes atypical adenomatous hyperplasia and adenocarcinomas in situ according to IASLC.

training set. We applied DL-feature_{initial} or DL-feature_{follow-up} into the ridge regression model to gain predictive value for the identification of GGNs. This predictive value was defined as DL-score. If a nodule had a higher DL score, it was more likely to be malignant.

To integrate the DL information mined from sequential CT images, we defined the following equation to gain the integrated DL-feature from different time points:

$$\text{DL-feature}_{\text{initial+follow-up}} = 1 - (1 - \text{DL-feature}_{\text{initial}}) \times (1 - \text{DL-feature}_{\text{follow-up}})$$

Then, we used DL-feature_{initial+follow-up} to build the ridge regression model to distinguish benign from malignant GGNs.

Since some clinical characteristics have shown the ability to estimate GGN risk,^{22–28} we evaluated nine clinical factors (sex, age, smoking, extrapulmonary cancer history, family cancer history, nodule size, type, location, and the time interval between two CT scans) and chose the significant predictors ($p < 0.05$) in the training set to construct the clinical model.

Considering that additional patient conditions can be described from clinical risk factors, we further incorporated clinical predictors into the DL-features to build the combined model for individualized prediction of GGNs.

Observer study

To compare the DL model with human performance, two radiologists (senior radiologist, ZC with 10+ years of experience; and junior radiologist, KS 3+ years of experience) were enrolled. They were blinded to the clinicopathological results to diagnose all GGNs in the test set. They first classified the GGNs based on the initial CT and then added follow-up CT scans for further diagnosis.

Statistical analysis

All statistical analyses were performed with R software (version 3.6). Significant differences were assessed by the chi-square test for categorical variables and the *t*-test for

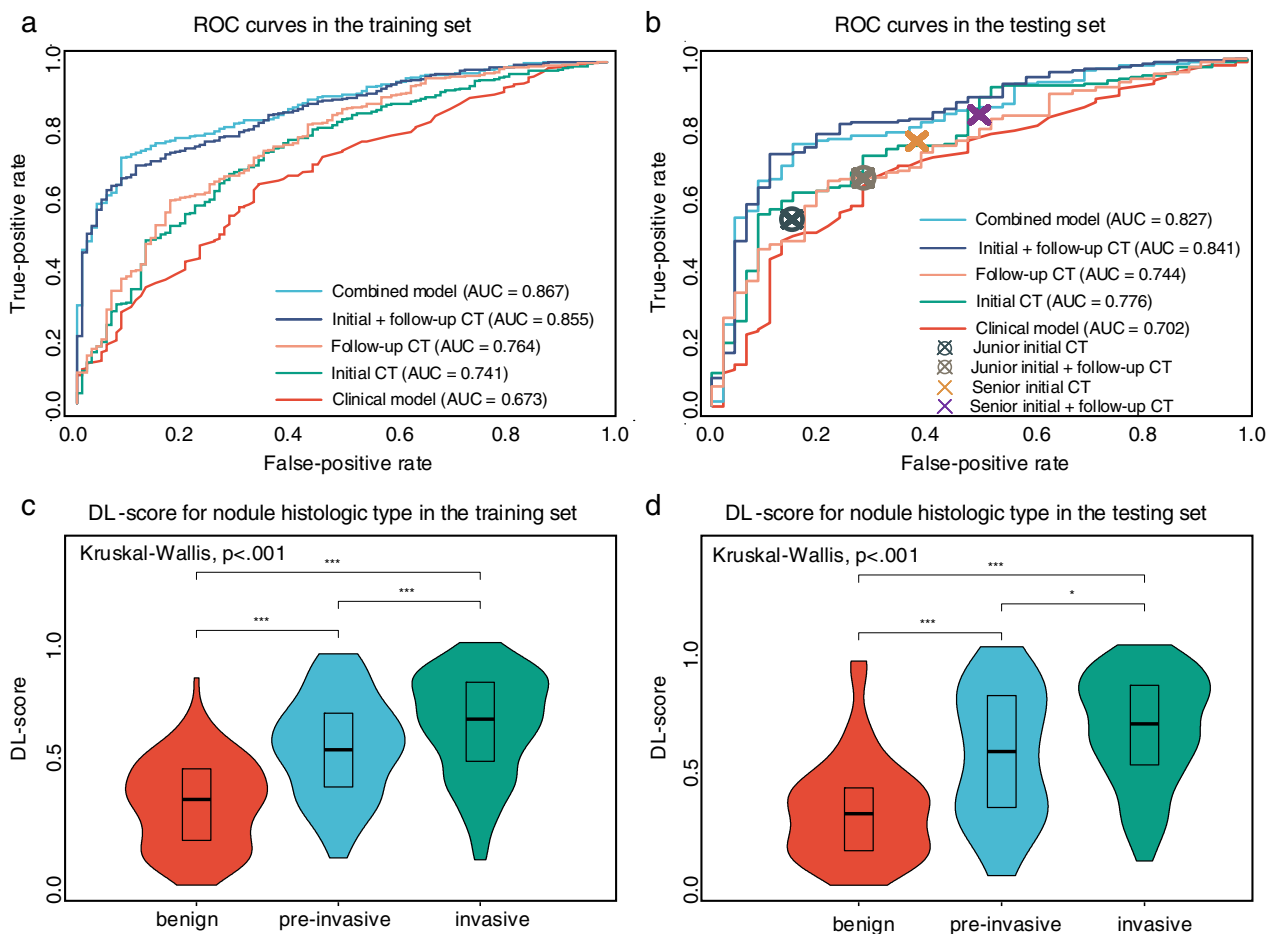


FIGURE 2 Performance of the various models. (a) Receiver operating characteristic curves of the clinical model, DL model using initial CT, DL model using follow-up CT, DL model using initial CT + follow-up CT, and the combined model using initial CT + follow-up CT + clinical in the training and testing sets. (b) Receiver operating characteristic curves of the various models and two radiologists. (c and d) DL-score from the initial CT + follow-up CT within the histological subgroup

T A B L E 2 Diagnostic performance of the models

	Training set				Testing set			
	AUC (95% CI)	ACC (%)	SEN (%)	SPE (%)	AUC (95%CI)	ACC (%)	SEN (%)	SPE (%)
Clinical model								
Age + extrapulmonary cancer history + nodule size	0.673 (0.617–0.729)	64.57 (60.23–68.73)	64.25 (59.34–68.95)	65.74 (55.99–74.60)	0.702 (0.619–0.784)	66.54 (60.37–72.31)	65.87 (58.99–72.80)	69.57 (54.25–82.60)
DL model								
Initial CT	0.741 (0.688–0.793)	68.11 (63.86–72.15)	67.25 (62.41–71.83)	71.30 (61.80–79.59)	0.776 (0.704–0.848)	64.96 (58.75–70.82)	60.58 (53.58–67.26)	84.78 (71.13–93.66)
Follow-up CT	0.764 (0.713–0.815)	64.57 (60.23–68.73)	60.50 (55.52–65.32)	79.63 (70.80–86.70)	0.744 (0.670–0.818)	67.32 (61.18–73.06)	64.90 (58.00–71.38)	78.26 (63.64–89.05)
Initial + follow-up CT	0.856 (0.821–0.890)	74.21 (70.18–77.96)	71.50 (66.80–75.88)	84.26 (76.00–90.55)	0.841 (0.777–0.904)	75.59 (69.83–80.74)	72.60 (66.00–78.54)	89.13 (76.43–96.38)
Combined model								
Initial + follow-up CT + age + extrapulmonary cancer history + nodule size	0.867 (0.834–0.900)	76.18 (72.23–79.82)	72.00 (67.32–76.35)	91.67 (84.77–96.12)	0.827 (0.763–0.891)	77.17 (71.50–82.18)	75.48 (69.05–81.17)	84.78 (71.13–93.66)
Junior radiologist								
Initial CT						60.14	53.81	85.00
Initial + follow-up CT						66.89	65.68	71.67
Senior radiologist								
Initial CT						73.31	76.27	61.67
Initial + follow-up CT						77.03	83.90	50.00

Abbreviations: AUC, area under the receiver operating characteristic curve; CI, confidence interval; ACC, accuracy; SEN, sensitivity; SPE, specificity.

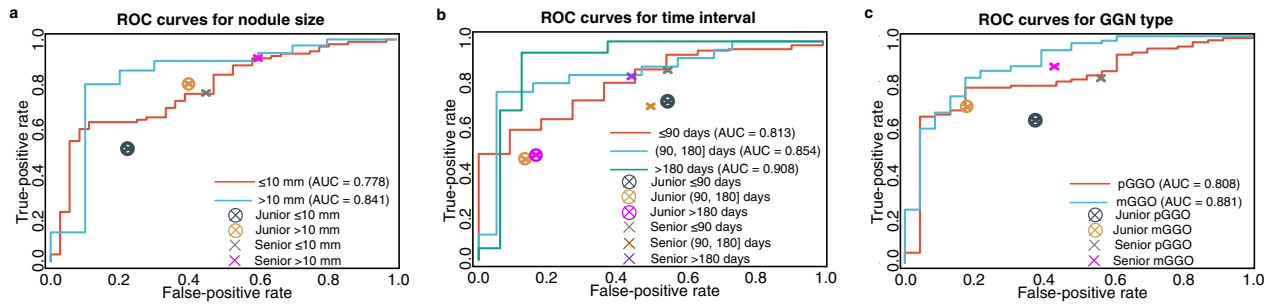


FIGURE 3 Performance of the DL model using initial and follow-up CT and radiologists within different subgroups in the testing set. (a) Receiver operating characteristic curves of the DL model and radiologists within tumor diameter subgroups. (b) Receiver operating characteristic curves of the DL model and radiologists within the time interval of two CT scanning subgroups. (c) Receiver operating characteristic curves of the DL model and radiologists within type subgroups

continuous variables. The difference in the area under the receiver operating characteristic curves (AUCs) between models was evaluated by the DeLong test. Youden index was used as the cutoff. $p < 0.05$ indicated a statistically significant difference.

RESULTS

Patient characteristics

Of the 762 patients enrolled in this study, the median age (SD) was 53.85 (11.4) years, 213 (28.0%) were men, and 103 (13.5%) were smokers. Patients in the training and testing sets were balanced for malignant prevalence (79% vs. 77.3%, respectively; $p = 0.35$). No significant differences were found in any of the clinicopathological characteristics between the training and testing sets (Table 1).

Diagnostic performance of models

In Figure 2b and Table 2, we compared the ROC curves generated by all models and the two radiologists. In the testing set, the clinical model yielded an AUC of 0.702 (95% CI: 0.619–0.784), the DL models using one time-point CT image yielded an AUC ranging from 0.744 to 0.776, and the DL model using two time-point CT images yielded significantly higher performance (AUC [95% CI] 0.841 (0.777–0.904) vs. 0.776 (0.704–0.848); $p < 0.05$). However, when combining the DL-feature_{initial+follow-up} with clinical factors (age, extrapulmonary cancer history, and nodule size), the combined model showed a slightly increased AUC of 0.867 in the training set and a slightly decreased AUC of 0.827 in the testing set.

The sensitivity of initial CT using the DL model, junior radiologist, and senior radiologist was 60.58, 53.81, and 76.27%, respectively. The sensitivity of initial and follow-up CT scans using the DL model, junior radiologist, and senior radiologist was 72.60, 65.68, and 83.90%, respectively. The specificity of DL models varied from 78.26% to 89.13%. In

contrast, the specificity of readers was 71.67%–85.00% for the junior radiologist and 50.00%–61.67% for the senior radiologist.

Stratified analysis

To relieve the impact of different nodule sizes, time intervals between two CTs, and GGN components, we conducted stratified analysis using two time-point CT images in the testing set (Figure 3 and Table 3). The DL model and radiologists both achieved higher diagnostic performance in the nodule size larger than 10 mm subgroup than in the subcentimeter subgroup (accuracy, 70% vs. 80% for the DL model; 57.69% vs. 77.14% for the junior radiologist; 70.51% vs. 84.29% for the senior radiologist). In the time interval subgroup, the results showed that, with a longer time interval between the two CT scans, the DL model achieved higher performance in distinguishing benign GGNs from malignant GGNs, and the AUC value increased from 0.813 (≤ 90 days) to 0.908 (> 180 days). In contrast, the junior radiologist showed consensus in different time interval subgroups, ranging in accuracy from 61.02% to 69.59%, while the senior radiologist showed variability in different time interval subgroups, ranging in accuracy from 62.71% to 82.47%. In the GGN-type subgroup, the DL model showed comparable performance within the pure GGN (pGGO) and mixed GGN (mGGO) subgroups, with AUCs ranging from 0.808 to 0.881. Meanwhile, the readers exhibited approximately 8% higher accuracy in the mGGO subgroup.

To further clarify the diagnostic ability of the DL model, we also compared the DL score from two time-point CTs within the pathologic subtype (Figure 2c and d). Since malignant nodules with preinvasive adenocarcinoma and invasive adenocarcinoma have different overall survival, treatment plans, and management, we compared the DL-score among benign, preinvasive, and invasive nodules. There were significant differences among these three pathological subtypes in both the training and testing sets (all $p < 0.05$).

TABLE 3 Diagnostic performance of the DL model and radiologists within different subgroups

	Training set						Testing set			Junior radiologist			Senior radiologist			
	AUC (95% CI)		ACC (%)	SEN (%)	SPE (%)		AUC (95% CI)		ACC (%)	SEN (%)	SPE (%)	ACC (%)		SEN (%)	SPE (%)	
Nodule size																
≤ 10 mm	0.831 (0.781–0.882)	75.10	74.44	76.71	88.90	0.778 (0.690–0.865)	70.00	62.80	57.69	50.86	77.50	70.51	75.86	55.00		
> 10 mm	0.872 (0.820–0.924)	81.57	80.45	88.57	90.00	0.841 (0.678–0.999)	80.65	79.82	77.14	80.00	60.00	84.29	91.67	40.00		
Time interval																
≤ 90 days	0.829 (0.775–0.883)	73.22	71.80	82.61	72.73	0.813 (0.703–0.923)	70.69	70.55	69.59	45.00	72.41	82.47	86.78	45.00		
(90, 180] days	0.834 (0.752–0.916)	74.19	72.06	80.00	94.74	0.854 (0.733–0.975)	84.44	76.92	61.02	45.95	86.36	62.71	70.27	50.00		
> 180 days	0.895 (0.815–0.975)	84.38	74.07	91.89	87.50	0.908 (0.785–0.999)	91.43	94.70	62.79	48.00	83.33	72.09	84.00	55.56		
Type																
pGGN	0.861 (0.814–0.909)	73.55	70.49	83.05	82.61	0.808 (0.714–0.903)	72.18	70.00	62.82	62.90	62.50	73.72	81.45	43.75		
mGGN	0.853 (0.804–0.903)	75.56	72.35	89.80	82.61	0.881 (0.801–0.962)	81.82	81.63	71.43	68.75	82.14	80.71	86.61	57.14		

Abbreviations: ACC, accuracy; AUC, area under the receiver operating characteristic curve; CI, confidence interval; mGGN, mixed ground-glass opacity; pGGN, pure ground-glass opacity; SEN, sensitivity; SPE, specificity.

To further interpret the DL model visually, we also depicted four representative prediction results in Figure 4. In this figure, it is difficult to distinguish benign GGN from malignant GGN only by clinical information or visual observation on CT due to the resembling clinicopathological characteristics. However, the DL model could yield discriminative predictive value. Moreover, we used *t*-distributed stochastic neighbor embedding (*t*-SNE) to reduce the 256-dimensional DL-feature into 2 dimensions (Fig. S2 in the Supplement).²⁹ We observed that benign GGNs were clustered away from malignant GGNs.

DISCUSSION

GGN, a common finding on chest CT scans, comprises a variety of disease categories, and the guidelines for managing GGN are distinct from those for managing solid nodules.^{30,31} The malignancy rate of persistent GGNs is also higher than that of solid nodules.³² Long-term follow-up CT is recommended for low-risk GGNs.^{33,34} Therefore, an automatic GGN malignancy prediction method from sequential CT scans can provide auxiliary but important medical insights. In this research, this clinical problem was addressed by integrating the deep learning features from baseline and follow-up CT images. This approach illustrated promising diagnostic performance on par with or better than radiologists in the observer study.

GGN is a dynamic biological system that is quite different from solid nodules.^{35,36} A high-risk indicator of GGN malignancy is a new solid component developed during follow-up.³⁷ Therefore, the characteristics of GGNs may not be completely captured on baseline CT images. Our study also illustrated that the DL model using two time-point CT images had significantly better performance than that using one time-point CT image (AUC (95% CI), 0.841 (0.777–0.904) vs. 0.776 (0.704–0.848); *p* < 0.05). This was consistent with the performance of radiologists. The results indicated that incorporating sequential CT images is key to capturing dynamic GGN changes and further identifying GGNs.

One strength of the DL method is that it can discover and learn abstract high-level features that are invisible to the human eye but can reflect the intrinsic characteristics of GGNs.³⁸ In contrast, radiologists diagnose GGN mainly based on typical radiographic features, such as nodule shape, size, border, component, and margin.³⁹ These qualitative features might be less specific to GGNs than the DL method. Interestingly, we found that the junior radiologist achieved higher specificity (71.67%) and lower sensitivity (65.58%), while the performance for the senior radiologist reversed this, with a specificity of 50.00% and sensitivity of 83.90%. In contrast, our DL method achieved a balanced sensitivity (72.60%) and specificity (89.13%). Furthermore, our DL method yielded diagnostic accuracy comparable with that of the senior radiologist and almost 9% higher accuracy than that of the junior radiologist (Table 2). Consequently, our DL method can therefore provide a helpful adjunct to radiologists.

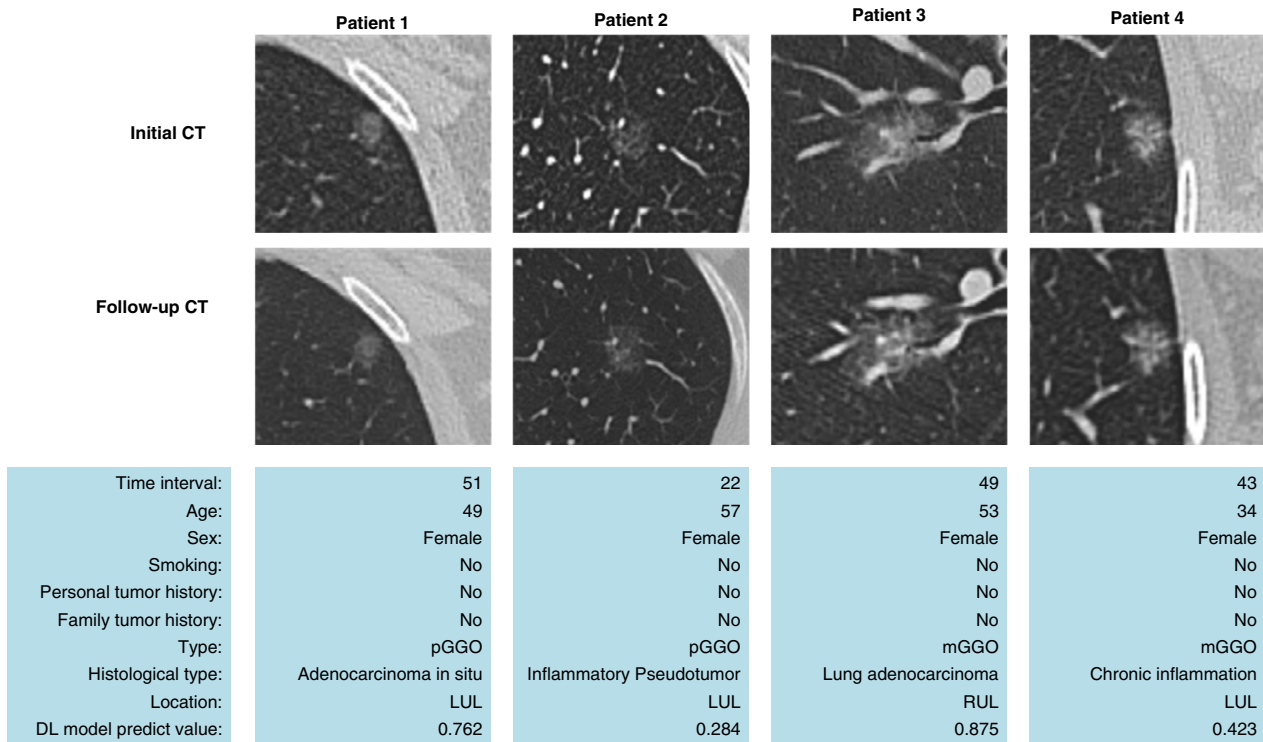


FIGURE 4 Representative prediction results from the testing set

Another strength of the DL method is that it does not need precise delineation of the lesion boundary. Due to the lower contrast of GGNs against the surrounding pulmonary parenchyma, it is difficult for clinicians to identify the boundary of GGNs. Previous studies using radiomic methods to diagnose GGN⁴⁰ might have been susceptible to inter-reader variability due to the precise contour of the lesion boundary.⁴¹

In previous studies, many clinical risk factors were found to be associated with malignancy in GGN, such as older age, extrapulmonary cancer history, nodule size, follow-up intervals, and the component of nodules.^{7,9,22,24,25,42} However, integrating clinical factors (age, extrapulmonary cancer history, and nodule size) with DL-features did not show an increase in performance. Additionally, we performed stratified analysis of the DL model and readers within nodule size, time intervals, and GGN type subgroups. As shown in Figure 3 and Table 3, the DL model achieved higher performance for nodule sizes larger than 10 mm. This was consistent with radiologists, indicating that subcentimeter GGN prediction is very difficult for both. Moreover, our results indicated that a longer time interval between the two CTs in the DL model led to higher performance in distinguishing benign GGNs from malignant GGNs. This can be explained by the fact that some malignant GGNs may progress and some benign GGNs may diminish in two time-point CT scans.³⁸ However, there were no general rules among radiologists within time interval subgroups. The junior radiologist maintained similar performance in different time interval subgroups, ranging in accuracy from 61.02% to 69.59%, and

the senior radiologist showed the best performance in the shortest time interval subgroup (≤ 90 days). The GGN component did not impact the performance of the DL model, with similar AUCs in the pGGN and mGGN subgroups in the training and testing sets. In contrast, the two radiologists both exhibited an approximately 8% accuracy drop in the pGGN subgroup. This decrease may be because the malignancy rate increases with increasing density of GGNs.⁴³

Among the malignant nodules, preinvasive adenocarcinoma and invasive adenocarcinoma have distinct overall survival, treatment planning, and management,⁴⁴ and many studies have used deep learning methods to predict tumor invasiveness from CT images.^{45–47} Therefore, we also compared the DL-score among benign, pre-, and invasive nodules to further interpret the diagnostic ability of the DL model using sequential CTs (Figure 2c and d). There were significant differences among these three pathological subtypes in both the training and testing sets (all $p < 0.05$). This indicated that, even though our DL model was not designed to predict tumor invasiveness, it could still distinguish preinvasive adenocarcinoma from invasive adenocarcinoma.

The first limitation of our research was selection bias due to the retrospective single-center dataset, and an external validation dataset is planned to generalize our DL model in the future. Second, deep learning cannot yield interpretable features for clinicians, as its overall process is similar to a “black box”. Therefore, we depicted four representative samples to interpret how the DL model functions.

In conclusion, building a DL-based model by sequential CT images is a cost-effective and noninvasive tracking

method to differentiate benign from malignant GGNs. Our DL method could achieve diagnostic performance on par with or better than radiologists to identify pulmonary GGNs.

ACKNOWLEDGMENTS

This study was supported by the Central Government Guided Local Science and Technology Free Exploration Projects of Sichuan Province, China (grant no. 2020ZYD005); National Natural Science Foundation of China (grant nos. 81700095, 91859203, 81871890, 81930053); China Postdoctoral Science Foundation (grant no. 2019TQ0019); the National Key R&D Program of China (grant no. 2017YFA0205200); Science and Technology Plan Project of Sichuan Province, China (grant no. 2019YFS0335).

CONFLICT OF INTEREST

The authors report no conflicts of interest.

ORCID

Zhixin Qiu  <https://orcid.org/0000-0003-0256-4419>

REFERENCES

- Liu Q, Liang W. The value of CT scan-guided biopsy in evaluation of ground-glass opacity. *Chest*. 2013;143:577. <https://doi.org/10.1378/chest.12-1950>
- Yabuuchi H, Matsuo Y, Tsukamoto H, Horiuchi T, Sunami S, Kamitani T, et al. Evaluation of the extent of ground-glass opacity on high-resolution CT in patients with interstitial pneumonia associated with systemic sclerosis: comparison between quantitative and qualitative analysis. *Clin Radiol*. 2014;69:758–64. <https://doi.org/10.1016/j.crad.2014.03.008>
- Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2021. *CA Cancer J Clin*. 2021;71:7–33. <https://doi.org/10.3322/caac.21654>
- Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011;365:395–409. <https://doi.org/10.1056/NEJMoa1102873>
- Church TR, Black WC, Aberle DR, Berg CD, Clingan KL, Duan F, et al. Results of initial low-dose computed tomographic screening for lung cancer. *N Engl J Med*. 2013;368:1980–91. <https://doi.org/10.1056/NEJMoa1209120>
- de Koning HJ, van der Aalst CM, de Jong PA, Scholten ET, Nackaerts K, Heuvelmans MA, et al. Reduced lung-cancer mortality with volume CT screening in a randomized trial. *N Engl J Med*. 2020;382(6):503–13. <https://doi.org/10.1056/NEJMoa1911793>
- MacMahon H, Naidich DP, Goo JM, Lee KS, Leung ANC, Mayo JR, et al. Guidelines for management of incidental pulmonary nodules detected on CT images: from the Fleischner society 2017. *Radiology*. 2017;284:228–43. <https://doi.org/10.1148/radiol.2017161659>
- Aberle DR, DeMello S, Berg CD, Black WC, Brewer B, Church TR, et al. Results of the two incidence screenings in the National Lung Screening Trial. *N Engl J Med*. 2013;369:920–31. <https://doi.org/10.1056/NEJMoa1208962>
- Kakinuma R, Noguchi M, Ashizawa K, Kuriyama K, Maeshima AM, Koizumi N, et al. Natural history of pulmonary subsolid nodules: a prospective multicenter study. *J Thorac Oncol*. 2016;11:1012–28. <https://doi.org/10.1016/j.jtho.2016.04.006>
- Henschke CI, Yankelevitz DF, Mirtcheva R, McGuinness G, McCauley D, Miettinen OS. CT screening for lung cancer: frequency and significance of part-solid and nonsolid nodules. *Am J Roentgenol*. 2002;178:1053–7. <https://doi.org/10.2214/ajr.178.5.1781053>
- Kobayashi Y, Fukui T, Ito S, Usami N, Hatooka S, Yatabe Y, et al. How long should small lung lesions of ground-glass opacity be followed? *J Thorac Oncol*. 2013;8:309–14. <https://doi.org/10.1097/JTO.0b013e31827e2435>
- Wiener DC, Wiener RS. Patient-centered, guideline-concordant discussion and management of pulmonary nodules. *Chest*. 2020;158:416–22. <https://doi.org/10.1016/j.chest.2020.02.007>
- Bi WL, Schabath MB, Mak RH, Abbosh C, Hoffmann U, Giger ML, et al. Artificial intelligence in cancer imaging: clinical challenges and applications. *CA Cancer J Clin*. 2019;0:1–31. <https://doi.org/10.3322/caac.21552>
- Kermany DS, Goldbaum M, Cai W, CCS V, Liang H, Baxter SL, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*. 2018;172:1122–1131.e9. <https://doi.org/10.1016/j.cell.2018.02.010>
- Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med*. 2019;25:954–61. <https://doi.org/10.1038/s41591-019-0447-x>
- Gao Y, Tan J, Liang Z, Li L, Huo Y. Improved computer-aided detection of pulmonary nodules via deep learning in the sinogram domain. *Visual Computing for Industry, Biomedicine, and Art*. 2019;2(1). <https://doi.org/10.1186/s42492-019-0029-2>
- Wang S, Zhou M, Liu Z, Liu Z, Gu D, Zang Y, et al. Central focused convolutional neural networks: developing a data-driven model for lung nodule segmentation. *Med Image Anal*. 2017;40:172–83.
- Wang S, Shi J, Ye Z, Dong D, Yu D, Zhou M, et al. Predicting EGFR mutation status in lung adenocarcinoma on computed tomography image using deep learning. *Eur Respir J*. 2019;53:1800986.
- Bankier AA, MacMahon H, Goo JM, Rubin GD, Schaefer-Prokop CM, Naidich DP. Recommendations for measuring pulmonary nodules at CT: a statement from the Fleischner society. *Radiology*. 2017;285:584–600. <https://doi.org/10.1148/radiol.2017162894>
- Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition 2017. p. 2261–2269.
- Deng J, Dong W, Socher R, Li L-J, Li K, Li F-F. ImageNet: a large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition; 2009. p. 248–255. [image-net.org/papers/imagenet_cvpr09.pdf](http://papers.image-net.org/papers/imagenet_cvpr09.pdf).
- Matsuguma H, Mori K, Nakahara R, Suzuki H, Kasai T, Kamiyama Y, et al. Characteristics of subsolid pulmonary nodules showing growth during follow-up with CT scanning. *Chest*. 2013;143:436–43. <https://doi.org/10.1378/chest.11-3306>
- Yang W, Sun Y, Fang W, Qian F, Ye J, Chen Q, et al. High-resolution computed tomography features distinguishing benign and malignant lesions manifesting as persistent solitary subsolid nodules. *Clin Lung Cancer*. 2018;19:e75–83. <https://doi.org/10.1016/j.clcc.2017.05.023>
- Dai C, Ren Y, Xie H, Jiang S, Fei K, Jiang G, et al. Clinical and radiological features of synchronous pure ground-glass nodules observed along with operable non-small cell lung cancer. *J Surg Oncol*. 2016;113:738–44. <https://doi.org/10.1002/jso.24235>
- Aokage K, Miyoshi T, Ishii G, Kusumoto M, Nomura S, Katsumata S, et al. Influence of ground glass opacity and the corresponding pathological findings on survival in patients with clinical stage I non-small cell lung cancer. *J Thorac Oncol*. 2018;13:533–42. <https://doi.org/10.1016/j.jtho.2017.11.129>
- Lee HW, Jin K-N, Lee J-K, Kim DK, Chung HS, Heo EY, et al. Long-term follow-up of ground-glass nodules after 5 years of stability. *J Thorac Oncol*. 2019;14:1370–7. <https://doi.org/10.1016/j.jtho.2019.05.005>
- Tammemagi M, Ritchie AJ, Atkar-Khattra S, Dougherty B, Sanghera C, Mayo JR, et al. Predicting malignancy risk of screen-detected lung nodules—mean diameter or volume. *J Thorac Oncol*. 2019;14:203–11. <https://doi.org/10.1016/j.jtho.2018.10.006>
- Chang B, Hwang JH, Choi Y-H, Chung MP, Kim H, Kwon OJ, et al. Natural history of pure ground-glass opacity lung nodules detected by low-dose CT scan. *Chest*. 2013;143:172–8. <https://doi.org/10.1378/chest.11-2501>

29. Cieslak MC, Castelfranco AM, Roncalli V, Lenz PH, Hartline DK. t-distributed stochastic neighbor embedding (t-SNE): a tool for eco-physiological transcriptomic analysis. *Mar Genomics*. 2020;51:100723. <https://doi.org/10.1016/j.margen.2019.100723>
30. Bai C, Choi C-M, Chu CM, Anantham D, Chung-Man Ho J, Khan AZ, et al. Evaluation of pulmonary nodules: clinical practice consensus guidelines for Asia. *Chest*. 2016;150:877–93. <https://doi.org/10.1016/j.chest.2016.02.650>
31. Baldwin DR, Callister MEJ. The British Thoracic Society guidelines on the investigation and management of pulmonary nodules. *Thorax*. 2015;70:794–8. <https://doi.org/10.1136/thoraxjnl-2015-207221>
32. Gao J-W, Rizzo S, Ma L-H, Qiu X-Y, Warth A, Seki N, et al. Pulmonary ground-glass opacity: computed tomography features, histopathology and molecular pathology. *Transl Lung Cancer Res*. 2017;6:68–75. <https://doi.org/10.21037/tlcr.2017.01.02>
33. Ost DE, Gould MK. Decision making in patients with pulmonary nodules. *Am J Respir Crit Care Med*. 2012;185:363–72. <https://doi.org/10.1164/rccm.201104-0679CI>
34. Silva M, Sverzellati N, Manna C, Negrini G, Marchianò A, Zompatori M, et al. Long-term surveillance of ground-glass nodules: evidence from the MILD trial. *J Thorac Oncol*. 2012;7:1541–6. <https://doi.org/10.1097/JTO.0b013e3182641bba>
35. Gould MK, Donington J, Lynch WR, Mazzone PJ, Midthun DE, Naidich DP, et al. Evaluation of individuals with pulmonary nodules: when is it lung cancer? Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest*. 2013;143:e93S–e120S. <https://doi.org/10.1378/chest.12-2351>
36. Jamal-Hanjani M, Wilson GA, McGranahan N, Birkbak NJ, Watkins TBK, Veeriah S, et al. Tracking the evolution of non-small-cell lung cancer. *N Engl J Med*. 2017;376:2109–21. <https://doi.org/10.1056/NEJMoa1616288>
37. Cho J, Kim ES, Kim SJ, Lee YJ, Park JS, Cho Y-J, et al. Long-term follow-up of small pulmonary ground-glass nodules stable for 3 years: implications of the proper follow-up period and risk factors for subsequent growth. *J Thorac Oncol*. 2016;11:1453–9. <https://doi.org/10.1016/j.jtho.2016.05.026>
38. Xu Y, Hosny A, Zeleznik R, Parmar C, Coroller T, Franco I, et al. Deep learning predicts lung cancer treatment response from serial medical imaging. *Clin Cancer Res*. 2019;25:3266–75. <https://doi.org/10.1158/1078-0432.CCR-18-2495>
39. S.M. Lee, C.M. Park, J.M. Goo, H.-J. Lee, J.Y. Wi, C.H. Kang, Invasive pulmonary adenocarcinomas versus preinvasive lesions appearing as ground-glass nodules: differentiation by using CT features., *Radiology* 268 (2013) 265–273. <https://doi.org/https://doi.org/10.1148/radiol.13120949>.
40. S.R. Digumarthy, A.M. Padole, S. Rastogi, M. Price, M.J. Mooradian, L. V Sequist, M.K. Kalra, Predicting malignant potential of subsolid nodules: can radiomics preempt longitudinal follow up CT?, *Cancer Imaging* 19 (2019) 36. <https://doi.org/https://doi.org/10.1186/s40644-019-0223-7>.
41. Liu Z, Wang S, Dong D, Wei J, Fang C, Zhou X, et al. The applications of Radiomics in precision diagnosis and treatment of oncology: opportunities and challenges. *Theranostics*. 2019;9:1303–22. <https://doi.org/10.7150/thno.30309>
42. Lee KH, Goo JM, Park SJ, Wi JY, Chung DH, Go H, et al. Correlation between the size of the solid component on thin-section CT and the invasive component on pathology in small lung adenocarcinomas manifesting as ground-glass nodules. *J Thorac Oncol*. 2014;9:74–82. <https://doi.org/10.1097/JTO.0000000000000019>
43. Lee SW, Leem C-S, Kim TJ, Lee KW, Chung J-H, Jheon S, et al. The long-term course of ground-glass opacities detected on thin-section computed tomography. *Respir Med*. 2013;107:904–10. <https://doi.org/10.1016/j.rmed.2013.02.014>
44. Travis WD, Brambilla E, Noguchi M, Nicholson AG, Geisinger KR, Yatabe Y, et al. International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma. *J Thorac Oncol*. 2011;6:244–85. <https://doi.org/10.1097/JTO.0b013e318206a221>
45. Xia X, Gong J, Hao W, Yang T, Lin Y, Wang S, et al. Comparison and fusion of deep learning and Radiomics features of ground-glass nodules to predict the invasiveness risk of stage-I lung adenocarcinomas in CT scan. *Front Oncol*. 2020;10:418. <https://doi.org/10.3389/fonc.2020.00418>
46. Gong J, Liu J, Hao W, Nie S, Zheng B, Wang S, et al. A deep residual learning network for predicting lung adenocarcinoma manifesting as ground-glass nodule on CT images. *Eur Radiol*. 2020;30:1847–55. <https://doi.org/10.1007/s00330-019-06533-w>
47. Zhao W, Yang J, Sun Y, Li C, Wu W, Jin L, et al. 3D deep learning from CT scans predicts tumor invasiveness of subcentimeter pulmonary adenocarcinomas. *Cancer Res* 2018;78(24):6881–9. <https://doi.org/10.1158/0008-5472.CAN-18-0696>.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Qiu Z, Wu Q, Wang S, Chen Z, Lin F, Zhou Y, et al. Development of a deep learning-based method to diagnose pulmonary ground-glass nodules by sequential computed tomography imaging. *Thorac Cancer*. 2022;13: 602–12. <https://doi.org/10.1111/1759-7714.14305>