Peer Review File

Article information: https://dx.doi.org/10.21037/jtd-23-985

Reviewer A

Comment 1: Since the term of medium-senior radiologist is confusing, please change as I mentioned. (eg. 10 and 15 years of experience in thoracic imaging)

Reply 1: Thank you very much for your kind comment. We have added this as advised (see Page 3, line 29; Page 3, line 46).

Changes in the text: "compared against radiologists with 10 and 15 years of experience in thoracic imaging (medium-senior seniority)" (see Page 3, line 29); "in comparison with that of the radiologists" (see Page 3, line 46).

Comment 2: In the introduction section authors said solid nodules are extremely common, with a malignancy rate of approximately 30%.

In subcentimeter sized solid nodules, malignancy rate is not high as 30%. If solid nodule is larger than 20mm, malignancy rate is high. Please find a relevant reference for this.

Reply 2: Thank you very much for your advice.

We have found some relevant references as your comment, and we are willing to present them to you. The probability of malignancy increases with the size of the SPN. As for subcentimeter solid nodules, the risk of cancer in nodules smaller than 6mm is estimated at less than 1.0%[1]; an estimated average risk of 0.5–2.0% is reported in solid nodules measuring 6–8mm[1, 2]; the risk is over 3.0% in nodules measuring greater than 8mm[2]. In two US studies, the prevalence of malignancy ranges from 2.3% to 6% for solid nodules 5–9 mm in diameter[3, 4]. However, the prevalence of malignancy in subcentimeter nodules reaches more than 20% in a Japanese study[5]. Godoy et al.[1] also proposed that the probability of malignancy is 5%-15% in solid nodules measuring 8mm–15mm.

Moreover, for solitary pulmonary nodule 11–20 mm in diameter, the prevalence of malignancy is 33–60%, and 64–82% in nodules greater than 20 mm[6]. Based on analysis of information from the Mayo Clinic CT Screening Trial, the probability of malignant pulmonary nodules larger than 20mm was 50%[7].

Here are the relevant references:

[1] M. Godoy, E. Odisio, M.T. Truong, et al. Pulmonary Nodule Management in Lung Cancer Screening: A Pictorial Review of Lung-RADS Version 1.0. Radiol. Clin. N. Am. 2018;56:353-363.

[2] MacMahon H, Naidich DP, Goo JM, et al. Guidelines for Management of Incidental Pulmonary Nodules Detected on CT Images: From the Fleischner Society 2017. Radiology 2017:284:228-243.

[3] Gohagan J, Marcus P, Fagerstrom R, et al. Baseline Findings of a Randomized Feasibility Trial of Lung Cancer Screening With Spiral CT Scan vs Chest Radiograph: the Lung Screening Study of the National Cancer Institute. Chest 2004;126:114-21.

[4] Henschke CI, Yankelevitz DF, Naidich DP, et al. CT Screening for Lung Cancer:

Suspiciousness of Nodules according to Size on Baseline Scans. Radiology 2004;231:164-8.
[5] Suzuki K, Nagai K, Yoshida J, et al. Video-assisted thoracoscopic surgery for small indeterminate pulmonary nodules: indications for preoperative marking. Chest 1999;115:563-8.
[6] Gould MK, Fletcher J, Iannettoni MD, et al. Evaluation of patients with pulmonary nodules: when is it lung cancer?: ACCP evidence-based clinical practice guidelines (2nd edition) Chest 2007; 132:108S-130S.
[7] Midthun DE, Swensen SJ, Jett JR et al. Evaluation of nodules detected by screening for lung cancer with low dose spiral computed tomography. Lung Cancer 2003; 41(suppl 2):S40.
Changes in the text: No changes in the text.

Comment 3: In the introduction section authors' mentioned Tang et al…. AUC of …, respectively.
Please mention for what sensitivity (DL model for what?).
Reply 3: Thank you very much for this question. We have added the explanation that the DL model was developed for classification of benign and malignant pulmonary nodules (see Page 7, line 83).
Changes in the text: "For instance, Tang et al. [19] proposed a DL model to improve the classification performance for benign and malignant pulmonary nodules, which…and 0.931, respectively" (see Page 7, line 83).

Comment 4: In clinical practice, it is difficult to characterize subcentimeter sized solid nodule for malignancy or benign due to the small size (difficult to determine is margin, border etc). Please included this in the limitation section.
Reply 4: Thank you very much for your advice. We have added this consideration to the limitations (see Page 15, line 262).
Changes in the text: "Besides, it is difficult to characterize SSPNs for malignancy or benign due to the small size in clinical practice, because some morphological features of SSPNs such as margin, border etc. are difficult to determine" (see Page 15, line 262).

Reviewer B

General comments: This article is about "Diagnostic Performance of a Deep Learning-Based Method in Differentiating Malignant from 2 Benign Subcentimeter Solid Pulmonary Nodule" The authors insisted that the DL-based method yielded higher performance in comparison with medium-senior radiologists in differentiating malignant and benign SSPNs. This DL model may reduce uncertainty in diagnosis and improve diagnostic accuracy, especially for SSPNs smaller than 8mm.

Strength: The advantage of this paper is that the already developed DL model for lung nodule diagnosis on chest CT was applied to the accuracy of lung cancer diagnosis in smaller lung nodules and showed good results compared to the results of radiologists.

Comment 1: Drawback: The disadvantage of this paper is that the number of cases applied to the study is too small, applying the existing developed lung cancer diagnostic DL model to lung nodules less than 1cm in size. It does not appear to be a scientific approach to mention these results for just 200 lung nodules less than 1cm. Further research is needed with at least 1000 lung nodules, and external validation is also required.

Reply 1: Thank you very much for your kind comment.

The deep learning approach proposed in this study is a mature and fully functional diagnostic model, which can already be used to assist radiologists to diagnose pulmonary nodules in clinical work. Training a new deep learning model requires a large amount of data, and at least 1000 cases of nodules are necessary. The model in our study was trained and validated based on a total of 5478 nodules (training dataset 4978; validation dataset 500) from five hospitals. The purpose of this study is to test whether the model is effective for subcentimeter solid nodules (SSPNs), rather than training a new model. The 200 SSPNs in this study served as an external validation dataset, and the testing results based on this dataset can also demonstrate the diagnostic performance of the model. To further improve the diagnostic efficiency of deep learning for SSPNs, we plan to develop a new diagnostic model specifically for SSPNs in further research and expect to collect and include 2000 nodules for training and validation.

Changes in the text: No changes in the text.

Comment 2: The biggest question is whether the radiologist's results can be trusted. I think it is difficult to agree with the results of the radiologist in Figure 5 presented by the authors, and perhaps other radiologists also find it difficult to agree if they see these pictures.

The two mentioned above are the deadliest shortcomings of this paper.

There are no comments for each section.

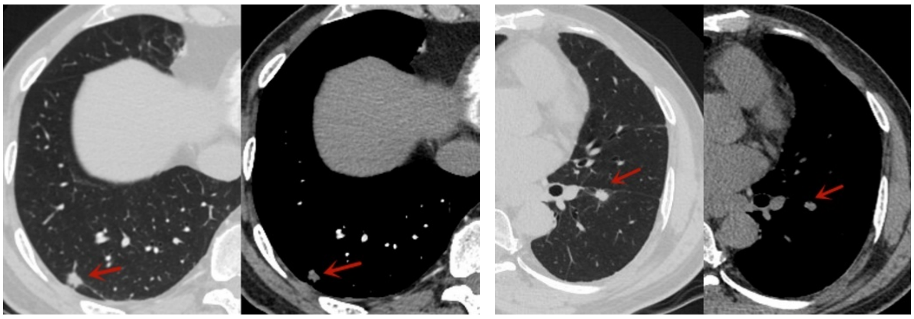
Reply 2: Thank you very much for your question.

We completely agree with you that it is important to provide some examples that can demonstrate the diagnostic level of the radiologists. For SSPNs, radiologists sometimes tend to give a relatively conservative diagnosis and advise patients to do a follow-up in our clinical practice, which cause some indeterminate results. We confirm that all the results in this study came from the radiology reports in clinical work and were not reviewed by chest imaging experts, representing the average diagnostic level in daily work. Several factors, such as overworking, lack of experience, and less discussion with other radiologists, may affect radiologists' performance. Thus, accuracy may be lower than when the cases are carefully reviewed by experts. We are willing to recheck the Figure 5 in this study and have replaced Figure 5c with a more representative image (see Page 22, line 400).

Changes in the text: "(C) A nodule in RLL of a 58-year-old female, with mean diameter of 6.5 mm…Histopathological finding: invasive adenocarcinoma" (Figure 5c, Figure Legends see Page 22, line 400)

Figures in the original manuscript：

a　　　　　　　　　　　　　　　　　b



c　　　　　　　　　　　　　　　　　d

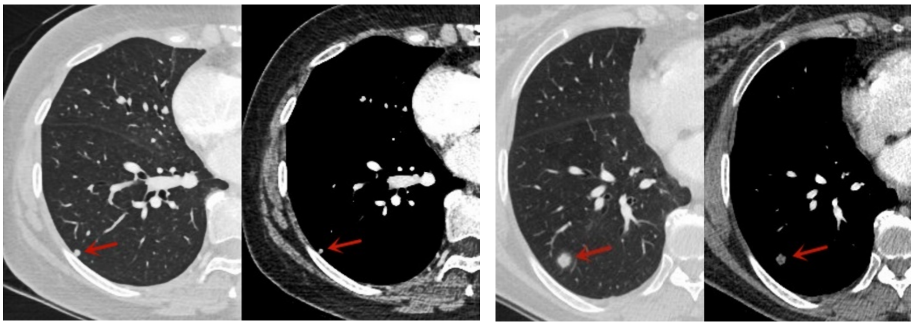
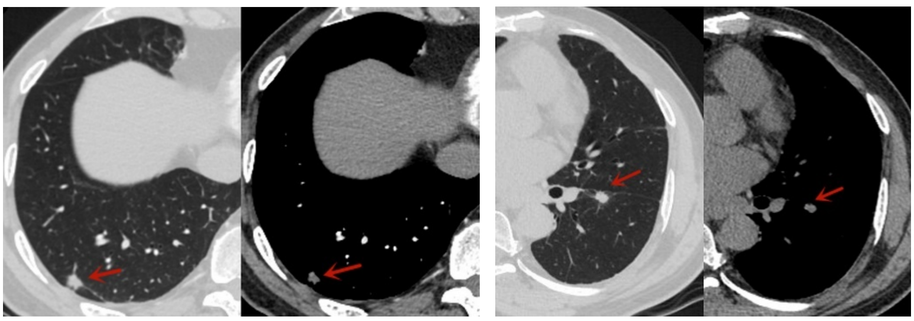


Fig. 5 Examples of nodules diagnosed by the AI model and the radiologists
(a) A nodule in RLL of a 46-year-old male, with mean diameter of 8.5 mm. The AI model diagnosed it as malignant, while the radiologists provided an indeterminate result. Histopathological finding: invasive adenocarcinoma. (b) A nodule in LLL of a 58-year-old male, with mean diameter of 8 mm. The AI model diagnosed it as benign, while the radiologists diagnosed it as malignant. Histopathological finding: pulmonary lymph node. (c) A nodule in RLL of a 57-year-old female, with mean diameter of 4 mm. The AI model diagnosed it as benign, while the radiologists gave an indeterminate result. Histopathological finding: pulmonary lymph node. (d) A nodule in RLL of a 53-year-old female, with mean diameter of 8.5 mm. The AI model diagnosed it as malignant, while the radiologists provided an indeterminate result. Histopathological finding: nonspecific inflammation.
AI, artificial intelligence; *RLL*, right lower lobe; *LLL*, left lower lobe

Figures in the revised manuscript：

a　　　　　　　　　　　　　　　　　b



c　　　　　　　　　　　　　　　　　d

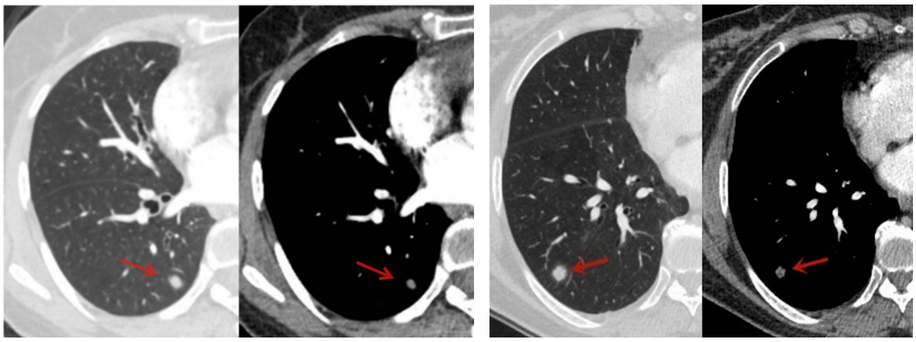


Fig. 5 Examples of nodules diagnosed by the AI model and the radiologists
(a) A nodule in RLL of a 46-year-old male, with mean diameter of 8.5 mm. The AI model diagnosed it as malignant, while the radiologists provided an indeterminate result. Histopathological finding: invasive adenocarcinoma. (b) A nodule in LLL of a 58-year-old male, with mean diameter of 8 mm. The AI model diagnosed it as benign, while the radiologists diagnosed it as malignant. Histopathological finding: pulmonary lymph node. (c) A nodule in RLL of a 58-year-old female, with mean diameter of 6.5 mm. The AI model diagnosed it as malignant, while the radiologists diagnosed it as malignant. Histopathological finding: invasive adenocarcinoma. (d) A nodule in RLL of a 53-year-old female, with mean diameter of 8.5 mm. The AI model diagnosed it as malignant, while the radiologists provided an indeterminate result. Histopathological finding: nonspecific inflammation.
AI, artificial intelligence; *RLL*, right lower lobe; *LLL*, left lower lobe