Peer review file

Article information: http://dx.doi.org/10.21037/jtd-20-2981

Reviewer A

Comment 1: For example, the peritumoral tissue is mentioned in the abstract, introduction and methods. This is not evaluated separately resp mentioned in the results. Rather, there are two types of evaluation: the isolated round lung tumor (GTV) and the lung tumor extended by the region of 10 voxel (GPTV). The pure PTV is never described as such in the results. Therefore, the abstract and the method should be adapted. Furthermore, it is not clear to me why the soft tissue window is also evaluated. This means that not only GGN but on subsolids nodes are included. If this is the case, this should be discriminated and mentioned. In addition, a subgroup analysis is recommended with regards of these two groups.

Reply: Thanks for this comment. We rewrote the conclusion part of the abstract and method to make it clearer that the trained deep learning models showed capabilities to predict GGN invasiveness. As we did not evaluate the effectiveness of the model, we removed this word accordingly. The objective of the evaluation of pulmonary nodules in soft tissue Windows is to determine whether pulmonary nodules are mGGNs or pGGNs. In this study, 63.32% of the nodules were pGGNs. The demographic and clinical characteristics of all patients are added to the text and Table 2. In clinical work, a considerable number of IAC show pGGNs in CT images. The data of this study showed that 108 of 311 patients with IAC showed pGGNs. The method of evaluating invasiveness based solely on CT features of the nodule will confuse the radiologists. The model of this study did not further divide the subsolid nodules into subgroups. Meanwhile, the sample size of the subgroup is relatively small, and the model is easy to over-fitting. The pathological invasiveness research of different sizes and densities is the direction of our future research with increasing sample size.

Comment 2: Since there are already many studies on this topic, the benefit regarding PTV is generally relatively spars and I wonder if there wouldn't be a connection with

simple parameters like the HU histogram.

Reply: Our prior study on the predictive the degree of pathological of HU histogram have been published (1). The results showed that CT value histogram can be used to predict the degree of pathological invasiveness. However, this method is time-consuming and poor clinical maneuverability. Therefore, the CT value histogram was not measured in this study. The deep learning methods can provide automatic and more comprehensive feature extraction works from the preoperative CT images.

1.Qiong L , Ya-Feng G , Li F , et al. Effect of CT window settings on size measurements of the solid component in subsolid nodules: evaluation of prediction efficacy of the degree of pathological malignancy in lung adenocarcinoma[J]. The British journal of radiology

Comment 3: It would also be interesting to see to what extent the selected peritumoral tissue range is relevant.

Reply: Thanks a lot for your comments. We mainly referred to the work done by Wu's groupand Beig's group[1-2], in which extending the VOIs by 5mm from the nodule boundary was used and showed satisfied performance. The pathological invasiveness research regarding peritumoral tissue range is the direction of our future research. Your comments can make our study more meaningful in the further.

1. Perinodular and Intranodular Radiomic Features on Lung CT Images Distinguish Adenocarcinomas from Granulomas.

2. Diagnosis of Invasive Lung Adenocarcinoma Based on Chest CT Radiomic Features of Part-Solid Pulmonary Nodules: A Multicenter Study.

All in all interesting work, which needs some adjustments.

Translated with www.DeepL.com/Translator (free version)

Reviewer B

This article focused on the application value of deep learning model based on CT images of tumor and peri-tumor regions in predicting the invasiveness of GGNs. The authors investigated thin-section CT of 622 patients with 687 pulmonary GGNs,

retrospectively. The two VOIs (gross tumor volume (GTV) and peritumoral volume (PTV)) were used as input to train model for invasiveness classification of GGNs, and the authors evaluated the accuracy of the classification model by calculating sensitivity, specificity, and AUC of the ROC curve. The performance of the classification model for distinguishing the invasiveness of GGNs was improved by adding the PTV to the GTV.

Comment 1: # P5. line 102-103, and Table 1.

The detail of CT scanners is not explained enough. For example, there are some Aquilions.

Reply: Thank you for your suggestion. The statement mentioned above was added in line 102-106, section "materials and methods":

Changes in the text:

All patients with GGNs were evaluated using one of five multidetector CT scanners (manufacturer: Toshiba, General Electric or Philips) with the following settings: tube voltage ,120 kVp; tube current,50–150 mA; image matrix ,512 × 512 pixels; and 0.5 second scanning duration. Images were reconstructed using the standard algorithm reconstruction with thicknesses of 0.625-1mm. All patients were scanned in the supine position and at full inspiration, ranging from thoracic entrance to lung base. The details of CT parameters were shown in Table 1.

Comment 2: #P5. line 106

The Images with slice thickness of 0.626-1mm were reconstructed in the main text. However, the slice thickness ranged 0.625-1mm in Table 1.

Reply: Thank you for pointing out the error. We have already corrected it.

Changes in the text:

Images were reconstructed using the standard algorithm reconstruction with thicknesses of 0.625-1mm.

Comment 3: # P.7 line 132

The reason why surrounding 10 pixels were determined for the PTV was not explained.

Response: Thanks a lot for your question. We apologize that the description about the peritumor region in the article was confusing and misleading. For increasing the peritumor region, we mainly referred to the work done by Wu's group[1]and Beig's group[2], in which extending the VOIs by 5mm from the nodule boundary was used and showed satisfied performance. In our manuscript, we wrote 10 pixels, as on a 2D scan, 10 pixels were extended on two directions in total. Thus, the VOIs were extended on three directions by five pixels. We rewrote related parts in the section "Method" in the main text to clarify this process. Two referred paper were:

1. Diagnosis of Invasive Lung Adenocarcinoma Based on Chest CT Radiomic Features of Part-Solid Pulmonary Nodules: A Multicenter Study.
2. Perinodular and Intranodular Radiomic Features on Lung CT Images Distinguish Adenocarcinomas from Granulomas.

Changes in the text:

The peritumoral region was generated by the morphology processing method called dilation operation, which is setting the increasing size and expanding the surrounding pixels of the original region within this size (shown in Figure 1). With reference to Wu et al. and Beig et al.'s work, the increasing size of the peritumor region was set to 5 pixels(22,23).

Comment 4: #P.7 line 132

How did the authors deal with the PTV for the nodules adjacent to the chest wall or mediastinum?

Response: Thanks for your question. If a VOI was extended to the chest wall or mediastium, the part out side of the lung region would be excluded according to the HU value. And then the VOIs would be processed in the same way of other VOIs.

Comment 5: #P.7 line 132

Were there any differences about the peritumoral region on CT between preinvasive and invasive GGNs? I think it would be important to interpret the utility of including peritumoral region of GGNs.

Reply: Thank you for your suggestions. We made a statistical analysis of the CT features of peritumoral region. The peritumoral region CT features included margin (smooth, lobular or spiculated), pleural indentation, nodule-lung interface (well defined or ill defined) and peritumoral region emphysema. There were no significant differences in peritumoral emphysema ($p = 0.072$). There were significant differences in margin ($p < 0.001$), nodule-lung interface ($p < 0.035$), pleural indentation ($p < 0.001$) between the two groups. The relevant data were summarized in Table E1.

Table E1 Relationship between Peritumoral CT features and pathological invasiveness of GGNs

Peritumoral CT Feature	**non-invasive lesions (n = 376)**	**invasive lesions (n = 311)**	**p**
Margin			<0.001
Smooth	349	205	
Lobular or spiculated	30	106	
Nodule-lung			0.035
interface	312	238	
Well defined	64	73	
Ill defined			<0.001
Pleural indentation	17	60	
Yes	359	251	
No			0.072
Peritumoral	14	21	
emphysema	362	290	
Yes			
No			

Comment 6: #P.7 line 136-137.

I would like to know the reconstruction kernel (for lung parenchyma or soft tissue structures) when evaluating the pulmonary nodules in mediastinal window.

Reply: In this study, mediastinal window was used to define the“solid” components of subsolid nodules. The visibility of the subsolid nodules was evaluated on routine mediastinal window (MW) settings (WW 300, WL 60), using a standard-resolution

algorithm. The soft tissue structures can be well displayed in this reconstruction kernel. The vessels in the subsolid nodules were excluded when evaluating the visibility of the subsolid nodules. We mainly referred to the preliminary work of our team [(1)].

1.Tu W T, Li Z B , Wang Y , et al. The "solid" component within subsolid nodules: imaging definition, display, and correlation with invasiveness of lung adenocarcinoma, a comparison of CT histograms and subjective evaluation[J]. European Radiology, 2018.

Comment 7: #P.7 line 143 and Table 2

The numbers of MIA and AAH are different between the main text and Table 2.

"… of which 115 were AAH, 113 were MIA, 148 were AIS, …" in the main text.

"AAH 113, MIA 115" in Table 2.

Reply: Thank you for your reminding of this mistake. In order to ensure that the data are true and correct，we rechecked the raw data and corrected the related contents in the main text and Table 2.

Changes in the text:

The data collected in this study included four types, of which 113 were AAH, 148 were AIS, 115were MIA, and 311 were IAC

Comment 8: #7. line 147

It would be better to present reference about the sentences "Patients diagnosed with AAH / AIS / MIA is feasible for sub-lobar resection, and the disease- free survival rate after surgery can be close to 100%, which is significantly better than invasive lung adenocarcinoma (38–86%, p <0.001).".

Reply: Thank you for your comments. We inserted the related references into the revised manuscript.

Changes in the text:

Patients diagnosed with AAH / AIS / MIA is feasible for sub-lobar resection, and the disease- free survival rate after surgery can be close to 100%, which is significantly better than invasive lung adenocarcinoma (38–86%, p <0.001)[(1)]

1.Yanagawa N , Shiono S , Abiko M , et al. New IASLC/ATS/ERS classification and invasive tumor size are predictive of disease recurrence in stage I lung adenocarcinoma.[J]. Journal of Thoracic Oncology, 2013, 8(5):612-618.

Comment 9: #Table 2.

311 nodules were diagnosed as IAC. IAC is usually seen as a part-solid or solid nodule/mass in lung window setting on CT. However, the number of mixed GGNs (part-solid nodules) was 252, and smaller than that of IAC. Were there IACs with CT findings of pGGNs?

Reply: Thank you for your comments. Yes, your understanding is correct. The data showed that 108 of 311 patients with IAC showed pGGNs.

Comment 10: # There are two types of parenthesis for reference like (25) and [25].

Reply: Thank you for your comments. We have revised it in the manuscript.

Changes in the text:

IAC was invasive lesions, and the type was defined as 0 (25).

Reviewer C

Comment 1: Can you elaborate what difference is to expected in the lung parenchyma of non- invasive and invasive tumors?

Reply: Thank you for your comments. First of all, Bicheng Zhang[(1)]found that the density of peritumoral lymphatic micro-vessel might provide reference for regional lymph node (LN) metastasis in non–small cell lung cancer (NSCLC). Niha Beig [(2)] found that radiomic features from intranodular and perinodular regions of nodules can distinguish non-small cell lung cancer adenocarcinomas from benign granulomas. Based on the comprehensive analysis of previous studies, we put forward our design ideas and research topics. We made a statistical analysis of the CT features related to the peritumoral region. CT features related to the peritumoral region included margin (smooth, lobular or spiculated), pleural indentation, nodule-lung interface (well defined or ill defined), peritumoral region emphysema. In the invasive lesions, the CT features

of lobular, spiculated, pleural indentation and ill-defined nodule-lung interface were significantly more than the non-invasive lesions($P < 0.01$). The result was presented in Table E1. The differences of these CT features provide a imaging evidence for us to further develop a deep learning model based on GPTV.

Peritumoral CT Feature	**non-invasive lesions (n = 376)**	**invasive lesions (n = 311)**	**p**
Margin			<0.001
Smooth	349	205	
Lobular or spiculated	30	106	
Nodule-lung interface			0.035
Well defined	312	238	
Ill defined	64	73	
Pleural indentation			<0.001
Yes	17	60	
No	359	251	
Peritumoral emphysema			0.072
Yes	14	21	
No	362	290	

1. B Z, G Y, Y Z, et al. M2-polarized tumor-associated macrophages are associated with poor prognoses resulting from accelerated lymphangiogenesis in lung adenocarcinoma. Clinics (Sao Paulo, Brazil) 2011;66:1879-86.

2. Beig N, Khorrami M, Alilou M, et al. Perinodular and Intranodular Radiomic Features on Lung CT Images Distinguish Adenocarcinomas from Granulomas. Radiology 2019;290:783-92.

Table E1 Relationship between Peritumoral CT features and pathological invasiveness of GGNs

Comment 2: Is there a difference between the peritumoral environment in non-sold and solid adenocarcinomas?

Reply: This study only focused on peritumoral environment around ground-glass nodule. Solid nodules have not been studied. Thank you for your suggestion and it would be the research orientation for our further study.

Comment 3: can you provide the c statistic of all the models? What improvement is expected after addition of PTV to GTV?

Response: Thanks a lot for this commend. We redrew the figure to show ROC of models of all folds, as shown in Figure 3. After adding PTV to GTV, the model showed better performance on the prediction of GGN invasiveness, as PTV is also an important sign of GGN invasiveness in clinic. That's why we used GPTV to train a model and compared it with the model trained using GTV.

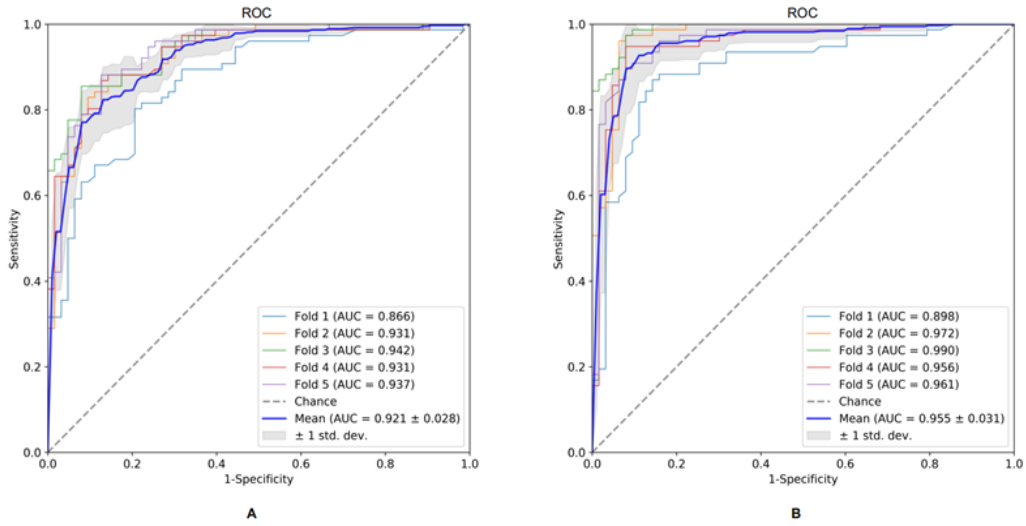

Figure3.The receiver operating characteristic (ROC) curves using two types of inputs (A. GTV B. GPTV).

Comment 4: The paper would benefit from an English editor, there are several grammatical errors.

Reply: We have followed your advice and polished the manuscript by professional expert.

Comment 5: Could you provide the size of solid and GGO component of all lesions? Could it be that the invasive tumors were larger than the non-invasive tumors and hence a larger PTV will be expected with invasive lesions hence such high prediction?

Reply: The research results of our team on the predictive the degree of pathological of solid components have been published (1). The results showed that 3D volume measurements of the solid component can be used to predict the degree of pathological malignancy. However, this method is time-consuming and poor clinical maneuverability. Therefore, the solid component was not measured in this study.

Your understanding is correct. From clinical work experience, the larger the tumor, the greater the probability of invasive adenocarcinoma. In order to study the ability of tumor and peritumoral size for predicting pathological invasiveness. The comparative experimental results show that tumor and peritumoral size can predict the invasiveness, and the AUC valued 0.837. However, using deep learning model, the AUC of the model increased to 0.955. The prediction accuracy of the fusion peri-tumor deep learning model is higher than that of the prediction efficiency through the size of GTV and PTV.

1.Qiong L , Ya-Feng G , Li F , et al. Effect of CT window settings on size measurements of the solid component in subsolid nodules: evaluation of prediction efficacy of the degree of pathological malignancy in lung adenocarcinoma[J]. The British journal of radiology

Comment 6: A supplemental table with patient and tumor demographics would be enhance the paper. How many patients had COPD. emphysema? How many were smokers etc.

Reply: Thank you for your suggestion, and your consideration is very reasonable.

The clinical characteristics of all patients have been added to Table 2.

Reviewer D

ABSTRACT

Comment 1: Line 23. It is not clear what is the “threshold value of 0.5”?

Reply: Thanks for your comment. We apologized for the confusion and mistake we made. This was a mistake as the trained models outputted the probabilities of IAC or not respectively and the class of higher probability was treated as the classification result of the case. We rewrote related part in abstract and methods.

Changes in the text:

The model using GTV showed a good performance in predicting GGN invasiveness, of which the AUC was 0.921 (95%CI, 0.896 - 0.937). Using GPTV, the AUC of the model increased to 0.955 (95%CI, 0.939 - 0.971).

Comment 2: Line 29. What is the definition of "efficiency" here?

Reply: Thanks for this comment. We rewrote the conclusion part of the abstract to make it clearer that the trained deep learning models showed capabilities to predict GGN invasiveness. As we did not evaluate the effectiveness of the model, we removed this word accordingly.

Changes in the text:

The method using deep learning obtained a good performance in the prediction of GGN invasiveness. The prediction ability of the model using GPTV was better than that of the model using GTV.

INTRODUCTION

Comment 3: 1st 2nd paragraph can be shortened.

Response:

Reply: Thanks for this comment. We have revised this part and made them more concise.

Comment 4: MATERIAL METHOD

Line124: it is not sure, how to determine the "largest lesion area"? for a mGGN, is it the largest total area or the solid area? Is it a single slice method or multiple slices to cover the whole nodules? If it is single slice method, cross tumor "volume" should be cross tumor "area"…

Reply: In this study, the maximal diameter of the tumor was used to evaluate the size of pulmonary nodules on axial CT images. The maximum diameter of the tumor was defined as the longest diameter, including GGO lesions. Multiple slices to cover the whole nodules as the gross tumor volume (GTV) was used for invasive assessment.

Comment 5: Line 130 why set the threshold = 10 pixel? Is it universal for different size of GGN? Why not proportional to GTV? This issue is critical

Reply: Thank you for your suggestion, and your consideration is very reasonable. First we need to apologize that the description about the peritumor region in the article was confusing and misleading. For increasing the peritumor region, we mainly referred to the work done by Wu's group and Beig's group[1,2], in which extending the VOIs by 5mm from the nodule boundary was used and showed satisfied performance. In the previous article, we wrote 10 pixels, as on a 2D scan, 10 pixels were extended on two directions in total. Thus, the VOIs were extended on three directions by 5 pixels. The 5 pixels are equivalent to 5mm evenly expanded around the tumor. Nodules of different sizes are expanded by 5mm. Until now, all current studies on the tumor peritumoral area are uniformly expanded. According to the GTV, equal proportion amplification is our directions for future research. Your comment can make our study more meaningful in the further.

We rewrote related parts in the main article to explain this process, please find in the Methods part. Two referred paper were:

1. Perinodular and Intranodular Radiomic Features on Lung CT Images Distinguish Adenocarcinomas from Granulomas.
2. Diagnosis of Invasive Lung Adenocarcinoma Based on Chest CT Radiomic Features of Part-Solid Pulmonary Nodules: A Multicenter Study.

Comment 6: RESULT

There is no external validation. Could Reviewer send a couple of cases to check the performance?

Reply: Okay, you're very welcome. It is necessary for external data sets to verify the

universality of our model. Our study is a single-institutional and retrospective study. Nevertheless, five CT scanners were used in our study and five-fold cross validation was performed. Finally, the deep learning model was found to be reproducible in the validation and training groups

Comment 7: It is interesting to do Sub-group analysis: e.g. < 1 cm or > 1 cm, pGGN or mGGN to see the difference

Reply: Thanks for your comment. The sample size of the subgroup is relatively small, and the model is easy to over-fitting. The pathological invasiveness research of different sizes and densities is the direction of our future research. This comment can make our study more meaningful in the further.

Comment 8:

DISCUSSION

Line 271 Why use automatically? It was manually done by one Radiologist and confirm by another. It is not automatically

Reply: Thank you for your suggestion, and your consideration is very reasonable. It is still a challenge to segment nodules by an automatic technique and all the ROI in the study were manually segmented. At present, our team is also engaged in automatic segmentation-related research. The automatic segmentation method will be used in the near future based on fully convolutional networks. The sentence was revised according to your suggestions as follows.

Changes in the text:

Our model could be used to accurately evaluate invasiveness of GGNs, improving the efficiency of lung cancer screening and assisting the radiologists to work efficiently. The automatic segmentation method will be used in the near future based on fully convolutional networks.

Comment 9: Limitation

Should mention NO external validation

Reply: Thanks for your comment.

The statement mentioned above was added in the in page 5-6 line 104-112, section "Discussion":

Changes in the text:

This is a single center study and lack of external validation set.

Comment 10: REFERENCE

Reference 33. 36 not complete

Reply: Thanks for your comment. We have made it complete.

Reviewer E

Methods

Comment 1: Please provide the size and proportion of the part-solid nodules of the AAH/AIS, MIA, and IAs, respectively, in Table 2.

Reply: Thank you for your comments. We have added the size and proportion information of GGNs in Table 2.

Comment 2: Page 7, line 144, "Table 2 is.. of patients." is repeated. It was already described in page 6, line 110.

Reply: Thank you for your comments. We have removed the repeated sentences.

Comment 3: Page 7, line 147, .. invasive lung adenocarcinoma (38-86%)" Please, add the reference for this sentence. P value of reference is not usually presented.

Reply: Thank you for your comments. We insert references to this part of the statement.

Comment 4:

Page 7, line 147-150, This sentence is also described repeatedly.

Reply: Thank you for your comments. We have removed the repeated sentences.

Comment 5: Page 7, line 151, "3d" means "3 dimensional"? Please, correct it.

Reply: Thank you for your comments. We have corrected the error here.

Comment 6: Page 8, line 163, Is this reference quoted correctly? It seems that this reference does not line with the previous sentence.

Reply: Thank you for your comments. We have corrected the error here.

Comment 7: There are redundancies of description in methods section. Please rewrite them concisely.

Reply: Thanks for your kind remind. We rewrote the methods part to make it more logical and legible.

Comment 8: Results

It would be better to move the first paragraph of the results section to the methods section. Instead, please add the description of the results of accuracy, sensitivey, specificity, PPV, and NPV in the results section.

Reply: Thanks for this comment. We rewrote the results section accordingly and added the description of accuracy, sensitivity, specificity, PPV and NPV.

Changes in the text:

Table 3 shows the performance of our CNN model using two types of inputs. According to the results, both using GTV and GPTV, the CNN model performs well. In each fold, comparing with using GTV, the performance of the model using GTPV all improved, though only Fold 3 existed statistically significant (DeLong's test: p=0.014). Overall, the mean AUC, accuracy, sensitivity, specificity, PPV and NPV of the model trained on GTV were 0.921 [95% CI: 0.896-0.937], 0.839 [95%CI: 0.812-0.868], 0.800 [95%CI: 0.759-0.843], 0.899 [95%CI: 0.851-0.924], 0.897 [95%CI: 0.863-0.927], and 0.787 [95%CI: 0.745-0.832] respectively, while same indexes of the model trained on GPTV were 0.955 [95%CI: 0.939-0.971], 0.904 [95%CI: 0.881-0.927], 0.893 [95%CI: 0.861-0.925], 0.917 [95%CI: 0.884-0.947], 0.929 [95%CI: 0.901-0.955], and 0.876 [95%CI: 0.841-0.912] respectively. The AUC differences between the deep learning-based model using GTV and using GPTV were statistically significant (DeLong's test:

p<0.01), indicating the accuracy of the classification model was improved by adding peritumor information. The corresponding ROC curves were shown in fig.3. Table 3 Note

Comment 9: Discussion

Page 10, line 206-208, Page 12, line 250-252, Two sentences are almost same. Please, delete one of them.

Reply: Thank you for your correction. We have removed the repeated sentences