Peer Review File

Article information: https://dx.doi.org/10.21037/jtd-2025-152

Reviewer A

Comment 1：R Studio is IDE not a package!

Reply 1: I sincerely apologize for the inaccurate statement in my manuscript. You are absolutely right that R Studio is an Integrated Development Environment (IDE), not an R package. This error was due to my momentary oversight. I have corrected this error in the manuscript.

Changes in the text: we have modified our text as advised (see Page 12, line 382, Reviewed Version of the Article)

Comment 2：They have not used proper metrics for the imbalanced dataset. See more in Interpretation of Artificial Intelligence Models in Healthcare: A Pictorial Guide for Clinicians.

Reply 2: Thank you for pointing out the concern about our use of metrics for the imbalanced dataset. We have taken steps to address this. We used a combination of ROC curves, calibration curves, and DCA for model evaluation. The AUC from the ROC curves offers reliable discrimination, as seen in our nodule analysis models. For data pre-processing, we have applied stratified random sampling to reduce bias, ensuring similar class distributions in training and validation sets. (We forgot to mention this grouping method in the Methods section.)

Changes in the text: We calculated sensitivity and specificity during internal analysis. In the Methods section, we added information about the methods for avoiding bias. (see Page 13, line 394-401, Page 19, line 592-612, Reviewed Version of the Article)

Comment 3：In addition, the database size was too small and imbalanced to be used for AI studies.

Reply 3: We fully agree with your opinion on the issues of database size and imbalance. In future research, we plan to actively collaborate with multiple centers to collect more data, thus expanding the database scale and improving its imbalance. At the same time, we will also use data augmentation techniques, such as oversampling the minority class data, to make the data distribution more balanced to meet the requirements of AI research.

Changes in the text: We added plans for addressing database issues in future research in the Discussion section (see Page 24, line 758-760, Reviewed Version of the Article).

Reviewer B

Comment 1：the study was based on retrospective single - center data with a small sample size, which may lead to selection bias and recall bias.

Reply 1: We are well aware of the issues you raised. Indeed, the retrospective single - center research design and the small sample size are limitations of this study, which may lead to selection bias and recall bias. During the research process, due to time and resource constraints, we chose this design. However, to reduce the impact of these biases, we strictly adhered to the established inclusion and exclusion criteria during data collection to ensure the representativeness of the sample. At the same time, we carefully cross - checked multiple data sources, such as medical records, imaging reports, and pathological results, to ensure the accuracy of the data and reduce recall bias. In the future, we plan to conduct multi - center prospective studies to expand the sample size, reduce the impact of biases on the research results, and improve the reliability and generality of the research conclusions.

Changes in the text: Added an elaboration on the research limitations and future improvement measures in the Discussion section (see Page 13, line 394-401, Page 24, line 758-760, Reviewed Version of the Article).

Comment 2：The manual delineation of ROIs lacked standardization, and no cases of lung squamous cell carcinoma were included. External validation through multi - center and large - scale studies is needed to confirm the reliability of the models.

Reply 2: We fully acknowledge the issues you have raised. Regarding the lack of standardization in the manual delineation of the region of interest (ROI), we are exploring the use of automated or semi-automated delineation tools. These tools have the potential to enhance the accuracy and standardization of delineation, minimizing the differences caused by human factors. However, it is undeniable that most current automated or semi-automated delineation methods are still not intelligent enough, and significant deviations often occur during the delineation process. Therefore, after referring to similar studies, we chose to have experienced radiologists perform manual delineation. In terms of case inclusion, in this study, the incidence of lung adenocarcinoma is much higher than that of lung squamous cell carcinoma. The number of squamous cell carcinoma cases is relatively small, and it is difficult to distinguish them from hilar structures on plain chest CT. Thus, such cases were not included. Nevertheless, in subsequent research, we plan to specifically include more lung squamous cell carcinoma cases and incorporate the delineation of enhanced CT scans of patients to ensure that the research covers different pathological types. In addition, we have already planned to conduct multi-center and large-scale studies. Through external validation, we aim to comprehensively evaluate the reliability of the models in different populations and clinical settings, thereby enhancing the generalizability of the research results.

Changes in the text: Supplementary content on ROI delineation improvement plans, case inclusion adjustment plans, and external validation was added to the Discussion

section (see Page 24, line 758-760, Reviewed Version of the Article)

Comment 3：In the title, I suggest the authors to directly indicate the development and validation of the diagnostic model.

Reply 3: We sincerely appreciate your insightful suggestion regarding the title. In response, we have made the necessary revisions to explicitly indicate the development and validation of the diagnostic model. We believe this adjustment will enhance the clarity and focus of our study title, making it more directly relevant to the core content of our research. Thank you again for your valuable input, which has undoubtedly contributed to the improvement of our manuscript.

Changes in the text: we have modified our text as advised (see Page 1, line 2-3, Reviewed Version of the Article)

Comment 4：The current abstract is not standardized and unnecessarily long. In the background, the authors did not analyze why radiomics-based model could accurately predict the benign versus malignant status and the degree of malignancy. In the methods the authors need to briefly describe the inclusion criteria, gold diagnoses of benign versus malignant status and the degree of malignancy, the generation of training and validation samples, and shorten the current text on the extraction of features and the model development. In the results, the authors did not briefly describe the sample characteristics and sensitivity and specificity of these models. The current conclusion has no detailed comments for the clinical implications of the findings.

Reply 4: We have comprehensively revised the abstract according to your suggestions. In the Background section, we added an analysis of why the radiomics - based model can accurately predict, explaining that radiomics features can reflect the characteristics of nodule tissues at the cellular level, enabling more accurate predictions. In the Methods section, we briefly described the inclusion criteria, the gold - standard diagnostic methods, and the generation of training and validation samples, and streamlined the content on feature extraction and model development. In the Results section, we added descriptions of sample characteristics and relevant data on the sensitivity and specificity of the models. In the Conclusion section, we elaborated on the guiding significance of the research results for clinical decision - making, such as helping doctors select appropriate treatment options.

Changes in the text: The abstract has been comprehensively revised (see Page 1, line 33- Page4 line 104, the Abstract section ,Reviewed Version of the Article)

Comment 5：In the introduction of the main text, the authors need to describe all available noninvasive diagnostic test methods for benign versus malignant PNs, have comments on their limitations and the potential strengths of models based on radiomics, and why the radiomics-based models are potentially accurate.

Reply 5: We have added relevant content to the Introduction section. We described traditional non - invasive diagnostic test methods in detail, such as imaging - based judgment methods based on nodule size, shape, density, etc., and tumor marker detection. We analyzed the limitations of these methods, such as the subjectivity of imaging feature judgment and the insufficient specificity of tumor marker detection. At the same time, we elaborated on the potential advantages of radiomics - based models, such as the ability to extract a large number of quantitative features and reflect the microscopic characteristics of nodules. We also explained the reasons for the potential accuracy of this model, that is, it can analyze nodules from multiple dimensions and comprehensively judge the benign or malignant nature of nodules.
Changes in the text: Added relevant content to the Introduction section (see page 5-6, line 125-155, Reviewed Version of the Article)

Comment 6：Because the authors used CM or COM models, they need to review the unique contrition of clinical features and why there is a need to consider clinical features to improve the accuracy of radiomics-based models.

Reply 6: We have added this content to the Discussion section. Clinical features, such as patients' age, gender, smoking history, etc., are closely related to the occurrence and development of pulmonary nodules. For example, patients with an older age and a long smoking history have a relatively higher risk of malignant nodules. Combining these clinical features with radiomics features allows for the evaluation of nodules from different perspectives. Clinical features provide the overall background information of patients, while radiomics features reflect the microscopic characteristics of nodules. The two complement each other, enabling a more comprehensive assessment of the benign or malignant nature and the degree of malignancy of nodules, thus improving the accuracy of the model.
Changes in the text: Added relevant content to the Discussion section (see page 19, line 619-634, Reviewed Version of the Article)

Comment 7：In the methodology, the authors need to correctly describe the primary clinical research and sample size estimation procedures of this study, as well as how benign versus malignant status and the degree of malignancy were ascertained or diagnosed.

Reply 7: We have revised the Methods section. We described in detail the process of the main clinical research, including the screening of patients and the specific steps of data collection. Regarding the sample size estimation, we explained the relevant studies we referred to and the estimation methods and basis we used. For the determination of the benign or malignant nature and the degree of malignancy of pulmonary nodules, we clearly stated that the postoperative pathological results were used as the gold - standard diagnostic method.
Changes in the text: Revised the relevant content in the Methods section (see page 8, line 222-240, Reviewed Version of the Article)

Comment 8：In statistics, please describe the P value for statistical significance, and the calculation of sensitivity and specificity since the two indicators are important for a diagnostic model.

Reply 8: We have added relevant content to the Statistics section. We clearly stated that in this study, a P - value less than 0.05 was considered statistically significant. However, to maximize the capturing of potential variables, seven variables with P values <0.2 were selected for multivariate logistic regression analysis. The minimum Akaike's information criterion (AIC) was utilized to select the optimal model parameters. We described in detail the calculation methods of sensitivity and specificity. Sensitivity = True positives / (True positives + False negatives), and Specificity = True negatives / (True negatives + False positives). We also supplemented the relevant calculation results in the Results section.

Changes in the text: Added relevant content to the Statistics section and the Results section (see page 13, line 394-401, Reviewed Version of the Article)

Comment 9：Please consider to cite several related papers: 1. Wu GF, Chen RC, Luo J, Li MT, Yu P, Shen PX, Luo JY, Qin YY. Diagnostic accuracy of folate receptor-positive circulating tumor cells in differentiating between benign and malignant pulmonary nodules. Transl Cancer Res 2024;13(12):6982-6994. doi: 10.21037/tcr-2024-2493. 2. Hou X, Wu M, Chen J, Zhang R, Wang Y, Zhang S, Yuan Z, Feng J, Xu L. Establishment and verification of a prediction model based on clinical characteristics and computed tomography radiomics parameters for distinguishing benign and malignant pulmonary nodules. J Thorac Dis 2024;16(3):1984-1995. doi: 10.21037/jtd-23-1400. 3.He XQ, Huang XT, Luo TY, Liu X, Li Q. The differential computed tomography features between small benign and malignant solid solitary pulmonary nodules with different sizes. Quant Imaging Med Surg 2024;14(2):1348-1358. doi: 10.21037/qims-23-995. 4. Liu J, Qi L, Wang Y, Li F, Chen J, Cheng S, Zhou Z, Yu Y, Wang J. Diagnostic performance of a deep learning-based method in differentiating malignant from benign subcentimeter (≤10 mm) solid pulmonary nodules. J Thorac Dis 2023;15(10):5475-5484. doi: 10.21037/jtd-23-985. 5. Xu S, Ge J, Liu X, He Q, Xu D, Cao W, Ding J, Kai X, Zhou G. The predictive value of chest computed tomography images, tumor markers, and metabolomics in the identification of benign and malignant pulmonary nodules. J Thorac Dis 2023;15(5):2668-2679. doi: 10.21037/jtd-23-250.

Reply 9: Thank you for your valuable suggestion to cite related papers. We have incorporated the five papers you recommended into our manuscript. These citations have enhanced our study by providing additional perspectives on diagnostic methods, model - building strategies, and radiological features, which strengthen the foundation and context of our research.

Changes in the text: we have Added these papers as advised

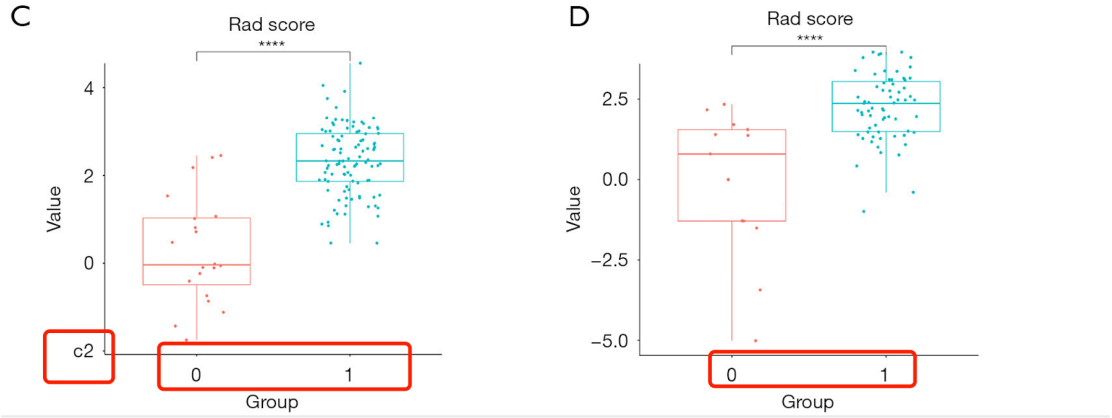
Comment 10：Some of the long sentences, especially in the "Discussion" section, could be rephrased to be more concise and easier to follow, but this is more of a matter of style preference rather than a strict language error.

Reply 10: Thank you for your feedback on the sentence length in the "Discussion" section. While we understand your concern, we believe that the current sentence structures are necessary to convey the complex ideas and arguments in this section. Each sentence has been carefully crafted to ensure scientific accuracy and comprehensiveness. We will, however, keep your suggestion in mind and will consider rephrasing if it does not affect the integrity of our research. Thank you again for your valuable input.

Reviewer B

1. Fig 4

1) Should “c2” be “-2”
2) Please provide the explanation of Group 0 and Group 1



Reply: I’ve revised fig4 and added the explanation of Group 0 and Group 1 in all the figures.

2. Fig 5, fig 14

Please provide the descriptions of the (also provide unit if applicable).

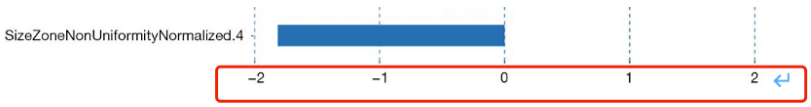


Figure 5 The importance of each feature included in the RM(I) in Study I. RM, radiomics model.

Reply: The X-axis represents impact coefficients of each feature, It is calculated through the statistical relationship of the data and reflects the relative change relationship between variables,

without specific physical units.

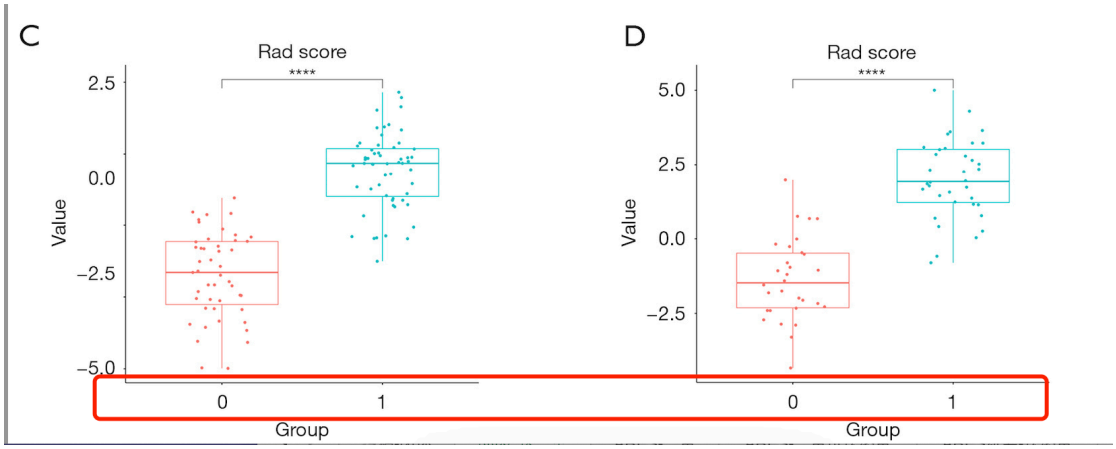
3. Fig 11, fig 17

Please provide the unit for age

Reply: I've revised fig11 and 17 as advised.

4. Fig 13

Please provide the explanation of Group 0 and Group 1.



Reply: I've revised fig13 and added the explanation of Group 0 and Group 1 in all the figures.

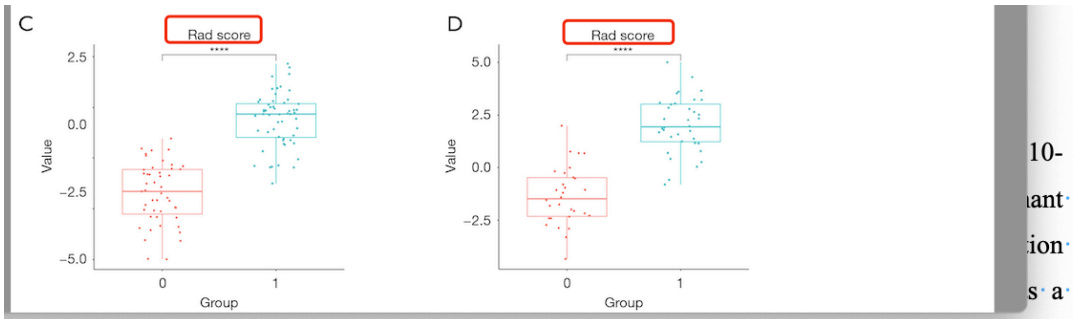
5. Fig 13D

Please indicate the legend for Fig 13D.

Reply: I'm sorry for missing the explanation of Fig 13D. It has now been added in the explanation.

6. Fig 4, Fig 6, Fig 8, Fig 13, fig 15

"Radscore" or "Rad score"? which one is correct? Please check the whole text and all figures and revise accordingly.



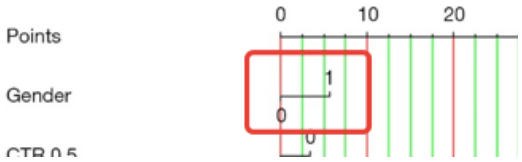
10-
ant
ion
s a

statistically significant difference in the distribution between the training group and the validation group ($p < 0.05$). Radscore, radiomics score; LASSO, least absolute

Reply: I've revised all the legend of the figures to be "Rad score".

7. Figure 8

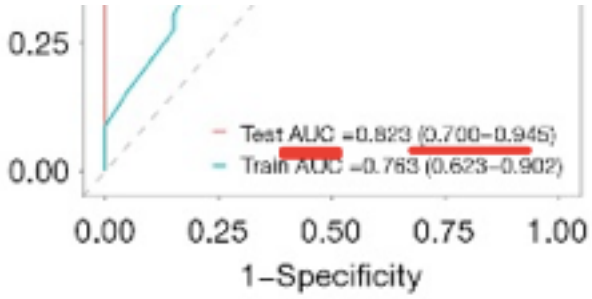
Please revise (0, 1) of gender to “Male, Female” in figure, or explain them in figure legend.



Reply：I have completed the revision of the legend of Figure 8,17 according to the requirements。

8. Figures 3A, 7A, 9A, 12A, 16A, 18A

(95% CI) should also be added. Please check and revise.



Reply：Referring to similar studies, I finally decided to add relevant explanations in the figure legend of Figure 3, 7, 9A, 12A, 16A, 18A to ensure its conciseness and clarity.