Peer Review File

Article information: https://dx.doi.org/10.21037/jtd-23-1400

Reviewer A

Comment: This is a retrospective evaluated radiomic data in addition to CEA and CYFRA21-1 and compared to pathological results. Authors showed a high sensitivity and specificity. It is not surprising that adding multiple tests together increases specificity. This combined approach has low risk to patients and can increase diagnostic accuracy.

Reply: Thank you very much for your recognition and guidance. All authors extend their heartfelt thanks to you. This study underscores the importance of integrating various diagnostic modalities and data sources to enhance clinical decision-making. It's a step forward in leveraging the advancements in radiomics alongside traditional markers, offering a more robust diagnostic framework for clinicians.

Reviewer B

Interesting and useful. I would like more details on the publication so that it can be reproducible.

Reply: Thank you for the feedback. You raise an important point-being able to reproduce published findings is a critical part of the scientific process. I agree more details on the methodology and results would be helpful here. In response to your valuable input, I will provide a more comprehensive overview of the methodology, ensuring that each step is clearly outlined to facilitate reproducibility. Additionally, I'll delve deeper into the results section, providing a more detailed analysis to enhance the clarity of our findings.

Reviewer C

1. Small sample size and number of lung cancers not reported.

Reply: Thank you for your valuable feedback. We appreciate your careful consideration of our work. In response to your comments regarding the sample size and the absence of reported cases of lung cancer, we would like to provide the following clarification: Sample Size Calculation PASS15.0 software was employed to process the sample size calculation. Based on previous studies, we estimated the effect size we hoped to detect and set power at 80% with an alpha of 0.05. This analysis suggested we would need a minimum of 97 subjects per group in order to have adequate power to test our hypotheses. We enrolled slightly more than this minimum number to account for potential attrition during the study. We have added it in the limitation.

2. No true external validation.

Reply：Thank you for your comments. We acknowledge that our model has not been externally validated. However, we have used cross-validation, bootstrapping etc. to get estimates of performance and generalization ability. We have added it in the limitation.

3. Lots of grammatical errors.

Reply: I understand your concern about grammatical errors. I'll be sure to thoroughly review the content and make necessary corrections. If you have specific examples or areas where you noticed issues, it would be extremely helpful if you could share them. This way, I can target those areas during the revision process.

4.AUC lower than achieved in other models.

Reply: Thank you for your insightful feedback regarding the AUC comparison. I understand the importance of benchmarking against other models. The observed lower AUC in our model can be attributed to several factors. Firstly, our dataset includes more diverse and challenging samples, which could contribute to a lower AUC but ensures robust performance in real-world scenarios. Additionally, our model architecture prioritizes simplicity to enhance interpretability, and this might impact AUC when compared to more complex models. Hyperparameter tuning and unique feature engineering further differentiate our approach. It's crucial to consider these

aspects in the context of the specific task requirements, where our model still demonstrates strong performance. I hope this clarifies the nuances of our model's AUC compared to others. Besides, the calibration curve and decision curve analysis concluded that the model is of great clinical value.

5. Nomograms are not as good to use as predictive models in a calculator.

Reply: Nomograms and predictive models in a calculator are two different ways of presenting clinical prediction models. Nomograms are graphical representations of statistical models that allow clinicians to calculate the probability of a clinical outcome for a patient based on the values of several predictors. On the other hand, predictive models in a calculator are computer programs that use statistical algorithms to predict the probability of a clinical outcome based on the values of several predictors. Both nomograms and predictive models in a calculator have their own strengths and limitations. While nomograms are easy to use and interpret, predictive models in a calculator can handle more complex models and can be used to make predictions for large datasets. We will take your feedback into consideration and work on improving our model.

6. There are a flood of pulmonary models now published and available that are better developed and tested with higher predictive accuracy.

Reply: Thank you for your comment. We appreciate your feedback. We acknowledge that there are many pulmonary models available that have been developed and tested with higher predictive accuracy. However, we believe that our model has its own unique strengths and can be useful in certain clinical settings. We will take your feedback into consideration and work on improving our model.

7. Few or no clinicians should use this model in practice.

Reply: Thank you for your comment. We appreciate your feedback. We acknowledge that our model may not be suitable for all clinicians. However, we believe that our model has the potential to benefit a specific group of clinicians and patients. Could you please elaborate on the reasons behind your statement? Understanding the specific concerns or limitations you have in mind would be valuable for evaluating the appropriateness of this model for clinical use. It's essential to address any potential

issues comprehensively to ensure the model's efficacy and safety in real-world settings. We will take your feedback into consideration and work on improving our model.