Peer Review File

Article information: https://dx.doi.org/10.21037/tlcr-23-666

Reviewer A

General comment: The article discusses the issue of predicting the growth of pGGNs, which can be a sign of early-stage lung cancers. The study introduces a DLCM that utilizes chest CT images to forecast the growth of pGGNs. The research was conducted on patients from two medical institutions and evaluated using various metrics such as accuracy, sensitivity, specificity, ROC curves, and AUROCs. The DLCM algorithm, DenseNet_DR, performed exceptionally well with an AUROC of 0.79 in predicting pGGN growth in the inner validation cohort and 0.70 in the external validation cohort. The introduction effectively emphasizes the significance of predicting pGGN growth and acknowledges the shortcomings of current methods.

Response: Thank you for your positive feedback. We are submitting a revised manuscript to address these concerns. Detailed point-by-point responses to these concerns are provided hereunder.

Specific Comments：

It would be helpful to include some statistics or references to back up the statement that the majority of growth in pGGNs occurs within 2-3 years of their detection.

Response:

Thanks for your advice. In the discussion section, we mentioned the growth of pGGNs occurs within 2-3 years in many previous studies. The discussion section are as follows “In this study, we chose 2 years of follow-up as the cutoff for defining stable pGGNs, which is one of the limitations of this study. Despite this, most growth of pGGNs occurs within 2 years (25.5–51.9%) or 3 years (41.0–86.0%) of its detection reported in previous studies (11,27,29-35).” Discussion section, paragraph4, page 9, line 286-290. The references were listed at reference section.

The discussion regarding the accuracy of the DLCM in various subgroups is quite informative. However, it would be even more beneficial if there were more elaborate explanations provided as to why certain subgroups exhibited superior prediction results.

Response:

Thanks for your valuable advice, we did find this interesting result and did a through discussion in the discussion section, paragraph 2 and 3, page 8-9, line 262-284. We demonstrated the reasons why shorter follow-up time subgroup and inner-validation subgroup shows superior prediction results. The reasons included CT image characteristics of different pGGNs, different follow-up time and different image conditions.

We think this were all the possible reason we can point out, however, there must be much more reasons beyond our thought. Could you please kindly review this part and point out is there any modification is needed, we appreciate your precious advice.

The discussion of limitations is comprehensive, recognizing the variability in the data, the absence of pathology confirmation, and the potential bias resulting from the limited follow-up period for stable pGGNs.

In conclusion, the summary highlights the main discoveries and recognizes the necessity for additional research and a more extensive dataset.

Overall, the article offers valuable insights into the application of deep learning models in predicting the growth of pGGNs. However, it would greatly benefit the analysis to include more context and references in certain sections and to further explain the factors contributing to the variations in model performance among different subgroups. Furthermore, it is crucial to consider the limitations and recognize potential sources of bias in order to maintain transparency in the study.

Furthermore, it is crucial to consider the limitations and recognize potential sources of bias in order to maintain transparency in the study.

Response:

Thank you for pointing this out, which we did not fully acknowledge in our original manuscript. We have added this in the conclusions section in the revised manuscript (pages 10, lines 322–323). "Considering all the limitations and the potential sources of bias, we believe further research and a larger database is required to validate and optimize our model."

Reviewer B

General comment: This paper highlights a Deep learning-based technique for pure ground-glass nodule prediction. Applying deep learning in the field of medicine is highly appreciated. This study is a relevant one and it falls within the scope of TLCR publication. However, the scientific content required some modification in a few places.

Response:

We appreciate the reviewer's encouragement and helpful comment. We are submitting a revised manuscript to address these concerns. Detailed point-by-point responses to these concerns are provided hereunder.

I recommend that the authors revise and resubmit the manuscript for further review. Some comments are given below:

Specific Comments:

1. Related works (Literature survey) can be included in the manuscript.

Response:

Thanks for your kind remind, we did a thorough search in the PubMed database, however, no such review which use artificial intelligence to predict pGGO growth. We will keep our knowledge updated and add the papers if we find them.

2. Why have you chosen 40x40 pixel areas of each nodule? Are there any technical reasons behind this? (Line 162)

Response:

Yes, it's for technical reasons. We choose 40*40 pixel areas because we not only need to include the detail of nodules as much as we could, but also within the capacity computing power of established model and computers' hardware. Before training our model, we preprocessed the samples. Each lung CT image is 512×512, and the average size of the labeled pulmonary nodules is about 30×30, which is too small in the image area. Therefore, we crop the original data for the nodular region at 30x30 and resize it to 40×40 as the data set for training and testing.

3. How many images (two classes- Growth pGGN and Stable pGGN) are available in the dataset?

The details can be given in tabulation format. 4. How many images are used for training and testing the deep learning models?

Response:

Thank you can carefully ask specific training details, because the training process is vital important to the final results. Question 3 and 4 are both related the training process, therefore we answered two questions together.

We totally enrolled 286 eligible patients, who had 419 pGGNs, including 253 persistent stable pGGNs and 166 with pGGNs that had grown. Because the pGGNs with varies size, then the images of each nodules we input in the training model was different. We totally in put 1107 grow pGGNs and 1102 stable pGGNs images in the training model, and 753 grow pGGN and 878 stable pGGN in the validation model. In totally, 3840 images with noduals are available in the dataset. Since we detailed introduced the clinic characteristics of patients and pGGNs, we hadn't fully described the detailed training process in this paper. Thank for the advises and we will add this data in the supplementary materials paragraph 3. In the paper, we add the following sentence "The detailed number of nodules images used in the training and validation set is provided in the supplementary material" at line 233-234 to further explain the images issue we used in the dataset.

5. For training the deep learning models, we need more images. How did you solve this issue during the training phase?

Response:

This is a particularly good question. We do have problem of less training images during the training phase. Since the original sample are not enough, we rotate the nodule image obtained after cutting, adjust saturation, brightness, contrast, sharpness and other data enhancement operations, in order to expand the size of the data set. Finally, the data set of this experiment has been expanded to: training set: 5,655 growth nodules and 5,184 stable nodules; validation set: 3774 growth nodules and 5004 stable nodules. We use the Dense-net for training and set batch size to 16, which means the network trains 16 samples each time. We set the initial learning rate to 1e-2 and apply the Adam optimizer to update the model parameters. The maximum iteration step is set to 2000.

6. Figure 3 can be explained in the Model development, evaluation and interpretability session. (Line 168 – 188).

Response:

Thanks for your advice. In line 168 – 188, we explained how we extract the nodules from each image, and the figure 3 do included how we did. However, figure 3 demonstrated how the heatmap and score-cam come from. For the model development, we descripted in supplementary materials. Since it's complicated and difficult to explanation just by words, we created the flow chat for the model development. The figure 1 and first paragraphs in supplementary materials tell the whole story of model development.

7. Why the performance of DenseNet_DR was better than other models? Technical Reasons must be included. (Line 234)

Response:

Thanks for your question, we compared the diagnose performance of four models, the AUC of other three were inferior to DenseNet-DR model and the difference was huge, therefore, we choose DenseNet_DR model as our model. As for the technical reasons, we detail descripted the mechanism and process in the supplementary material (paragraph 1 and 2) since it's huge space in the paper. We believe it's the technical reasons why dense-net DR model shows the best performance.

8. The performance metrics of the research work can be tabulated.

Response:

Yes, thanks for your kind advice. In table 3 (page 22), we detailed demonstrated the sensitive(sen), specific (SP), AUC, PPV, Positive predictive value; NPV, Negative predictive value.

9. The inference of all the figures shall be included in the discussion section.

Response:

Yes, thanks for your kind remind. We discussed the difference of all groups and subgroups at discussion secretion second and third paragraphs (Line 254-276). We also discussed the possible

reasons of the diagnose difference, such as image different image standard, follow-up time, cohort size and so on.

10. The results of this research work shall be compared with the related works and the significance of the proposed methodology should be highlighted.

Response:

Thanks for this valuable advises, we have tried to compare the related work previous studies. However, most previous have used quantitative CT imaging, such as the size, density, and volume of GGNs to estimated their invasiveness or used AI to estimate the mets of lyphm node. No such studies focus on predict the growth of GGNs, that's why we haven't do the comparation of related works. We sorry for that.