Peer Review File

Article Information: https://dx.doi.org/10.21037/tlcr-22-904

Review Comments (Round 1)

Reviewer A

General comments:

Structure and Syntax:

Review of the written language is recommended, especially:

o Improvement of orthography, ex. p.1 line 24 “prognostication in” – better: prognosis of

o Shorten long and complex sentences

o Ensure that the tempus is adequate and consistent throughout the whole manuscript

o Readability will be improved when repetition of words is avoided by using synonymous

→ We deeply appreciate your review of our manuscript and giving us a chance for revision. We followed your suggestion, and our manuscript was edited again by a native, English speaker, keeping these suggestions in mind.

Please underline the different approaches by exact quantitative research results. Following questions should be answered precisely for the presented theses:

2. How important is this specific factor for prognosis/ How big is the effect on negative or positive outcome? Eg. “strong prognostic factor”, “overall survival was longer” or “better survival” should be underlined by quantitative numbers to improve the value of these information

→ Thank you for your suggestion. In response to this valuable comment, we added quantitative numbers such as hazard ratios or odds ratios for each study result throughout the manuscript.

Changes in the text: All changes are noted with the track-change function, and we added comments to point out each change in the text (e.g., Reviewer A-2).

3. Finally, how big is the prognostic benefit expected to be by this new approach?

→ The magnitude of the survival benefit cannot be measured and is thus not reported in most retrospective studies. Thus, we could not summarize the magnitude of the survival benefits for the reported prognostic factors.

4. Concerning specific image analysis tools/ AI-pipelines: How good is the method for evaluation of the prognostic factor (statistical tests/analyses, DICE score, sensitivity/specificity etc.)?

→ The evaluation methods or metrics are beyond the scope of this mini-review.

5. For a better understanding and improving readability of the content supportive figures are recommended. E.g. CT examples of tumor measurements, of different tumor entities, morphologies and invasion (GGO, solid, margin, spiculae…), labeling/ segmentations of deep learning-based tools...

→ Thank you for your suggestion, but we aimed at briefly describing tumor-associated prognostic factors. We think that a text-based description would be sufficient for the purpose of this mini-review.

6. It is highly suggested to concern additional references including various authors, leading scientists and promising, recent studies worldwide – besides self-citation (~ 37%)!

→ We absolutely agree with your opinion. We additionally cited recent studies that reported valuable results for AI model-based prognostication (references #32-38 and #41-42) and review articles on CT-based prognostic features (references #1-3).

Changes in the text:

Page 3, line 3-6

Introduction

The prognostic importance of CT features, including the tumor dimension (1, 2), density (2), and radiomics features (3), has been summarized previously. Deep learning-based survival probability estimation and feature extraction are also feasible.

Page 9, line 7-23

Deep learning-based prognostic CT features

Deep learning-based prognostication is also feasible in advanced-stage lung cancers. Deng et al. (35) recently proposed a deep learning model using pretherapy CT scans, which provided a probability score to identify low-risk or high-risk patients receiving tyrosine kinase inhibitors and immune checkpoint inhibitors. The model was able to identify which patients would receive additional survival benefits beyond the median progression-free survival.

Deep learning is not limited to obtaining the cumulative survival probability; this method can be applied to capture the semantic features of lung cancer more broadly. Ahn et al. (41) reported that automatic measurements of the solid portions of lung cancer were comparable with manual measurements made by radiologists (intraclass correlation coefficient, 0.82-0.89) and showed good agreement with the invasive component size on microscopic examinations (intraclass correlation coefficient, 0.67). Kawaguchi et al. (42) showed that the solid component volume measured using deep learning had higher prognostic discrimination for recurrence or death (AUC, 0.752) than the solid component size (AUC, 0.722) or traditional three-dimensional volumetric analysis (AUC, 0.723). The presence of visceral pleural invasion was also estimated using

preoperative CT scans (31). In a study by Choi et al. (31), the diagnostic performance of the deep learning model (AUC, 0.75) for visceral pleural invasion was on par with that of board-certified radiologists (AUC, 0.73-0.79).

Specific comments:

Introduction:

7. Since only Tumor-associated prognostic factors are reviewed, please clarify the context and scope of the review at the end of the introduction, where two opposing information followed one by the other may be unclear for the reader:

1. P. 2 l. 43 f. “However, more information than tumor location and dimension can be acquired for the prognostication of lung cancers.”

2. P. 2 l. 45 f. “This mini-review briefly introduces tumor-associated prognostic factors that are extractable from chest CT scans”

→ We agree with your comment. We clarified the context and scope of our review in the Introduction (page 3, line 2-8).

Changes in the text:

Page 3, line 2-8

Introduction

Chest computed tomography (CT) scans are an essential tool for preoperative evaluations, clinical staging, and response monitoring in patients with lung cancer. The prognostic importance of CT features, including the tumor dimension (1, 2), density (2), and radiomics features (3), has been summarized previously. Deep learning-based survival probability estimation and feature extraction are also feasible. This mini-review briefly introduces both qualitative and quantitative tumor-associated prognostic factors that are extractable from chest CT scans.

Deep learning- based prognostic CT features & Other:

8. Recently there have been made several different approaches to acquire AI-based information from CT images, not only concerning radiomics/ radiogenomics, but also in improving quantitative and qualitative assessment of the tumor itself. To improve the quality of the manuscript integration of AI approaches for evaluation of the different prognostic, tumor-associated factors (location, spiculae, volume, etc.) is recommended. Further, concerning P. 3 l. 131 ff. “Nevertheless, CT definitions of pleural tag or retraction are highly variable among radiologists and often yield false-positive diagnoses for pathological visceral pleural invasion” a commend on possible AI solutions for this problem and the corresponding improvement of sensitivity/specificity is advised.
→ Thank you very much for your comment. We described more studies on deep learning-based lung cancer prognostication. Specifically, we described the following studies: 1) a model that could

provide a prognostic score for the additional survival benefit from tyrosine kinase inhibitors and immune checkpoint inhibitors among stage IV lung cancers; 2) a model that could measure the solid portion volume in early-stage lung cancers, which showed high agreement with manual measurements by radiologists; 3) deep learning-based measurements of the solid portion volume, which showed higher prognostic discrimination than manual diameter measurements and traditional volumetric analysis. In addition, we presented an AI algorithm for the prediction of visceral pleural invasion that showed radiologist-level diagnostic performance.

Changes in the text:

Page 9, line 7-23

Deep learning-based prognostic CT features

Deep learning-based prognostication is also feasible in advanced-stage lung cancers. Deng et al. (35) recently proposed a deep learning model using pretherapy CT scans, which provided a probability score to identify low-risk or high-risk patients receiving tyrosine kinase inhibitors and immune checkpoint inhibitors. The model was able to identify which patients would receive additional survival benefits beyond the median progression-free survival.

Deep learning is not limited to obtaining the cumulative survival probability; this method can be applied to capture the semantic features of lung cancer more broadly. Ahn et al. (41) reported that automatic measurements of the solid portions of lung cancer were comparable with manual measurements made by radiologists (intraclass correlation coefficient, 0.82-0.89) and showed good agreement with the invasive component size on microscopic examinations (intraclass correlation coefficient, 0.67). Kawaguchi et al. (42) showed that the solid component volume measured using deep learning had higher prognostic discrimination for recurrence or death (AUC, 0.752) than the solid component size (AUC, 0.722) or traditional three-dimensional volumetric analysis (AUC, 0.723). The presence of visceral pleural invasion was also estimated using preoperative CT scans (31). In a study by Choi et al. (31), the diagnostic performance of the deep learning model (AUC, 0.75) for visceral pleural invasion was on par with that of board-certified radiologists (AUC, 0.73-0.79).

Conclusion:

9. There are several algorithm-based commercial, open source and scientific tools already available, not only for quantitative and qualitative tumor assessment but also for comorbidities (e.g. YACTA for emphysema-quantification) - hence, the conclusion should be updated to the current state of the art (P5 l. 192 ff.” Future research should focus on the automated extraction of quantitative, tumor-associated and non-tumor-associated factors from CT scans and on the integration of this information with the staging system and clinical workflow.”). Especially since non-tumor-related prognostic factors were not part of the review, the exact separation in the conclusion and whole manuscript is suggested. Additionally, gene-panel testing (p.5 l.195ff.) was not part of the review

– In the end, a focus of the main goal of the manuscript, concluding the possibilities of CT image-based analysis of tumor-associated prognostic factors, is recommended.

→ Thank you very much for your comment. We agree with you that the conclusion should be succinct and focus on tumor-related prognostic factors. We revised the conclusion paragraph as you suggested, and all statements about non-tumor related factors or gene panel tests were removed throughout the manuscript.

Changes in the text:

Page 10, line-2-7

CT scans include rich prognostic information that could be extracted either by manual human interpretation or machine learning algorithms. In addition to measurements of tumor dimensions and T categorization in the staging system, CT images can be used to assess tumor volume, density, morphology, margin characteristics, and tumor-based survival probability. Future research should focus on the automated extraction and integrative modeling of these tumor-related prognostic factors.

Reviewer B

In their manuscript the authors present a mini-review on prognostic CT features in patients with lung cancer, covering both 'classical' features such as tumor size, as well as machine learning-based prediction models. The manuscript reads well and provides an interesting mix of tumor physiology and how this translates in to imaging, especially when it comes to tumor margin characteristics.

As this is a mini-review and not e.g. a systematic review, it is hard to judge the manuscript against clear set standards. My comments regarding this manuscript are centered around the liberal self-citations of the authors (11 out of 30 total citations) and the lack of citing other relevant research, in particular prior systematic reviews and other studies on deep learning.

Major comments:

1. The mini-review is not positioned against prior (systematic) reviews, nor are these prior reviews cited. Even though this is a mini-review, it should be put in perspective against other ¬reviews. E.g are prior reviews too broad? Or too narrow? Or too detailed? Or outdated? What is the added value of this review? Explaining what this mini-review adds compared to e.g. [1] and [2] would increase the relevance from a scientific point of view

→ We appreciate that you took the time to review this manuscript and your comments on our study.

The prognostic value of the tumor dimensional measurement and other semantic CT features has been summarized in the review articles. We briefly reviewed those features once again and added some more recent findings on central tumor location, ground-glass opacity, and deep learning-based survival probability estimation. As you suggested, the past review articles were cited in the Introduction.

Changes in the text:

Page 3, line 2-8

Introduction

Chest computed tomography (CT) scans are an essential tool for preoperative evaluations, clinical staging, and response monitoring in patients with lung cancer. The prognostic importance of CT features, including the tumor dimension (1, 2), density (2), and radiomics features (3), has been summarized previously. Deep learning-based survival probability estimation and feature extraction are also feasible. This mini-review briefly introduces both qualitative and quantitative tumor-associated prognostic factors that are extractable from chest CT scans.

2. The authors are liberal in citing prior work of their own. Specifically, 11 out of 30 citations include one or two of the authors as co-authors (namely references 5, 6, 7, 20, 21, 23, 26, 27, 29, 30). While self-citing is not a problem in itself, and this is a mini-review so there are no clear standards regarding how references should be sought, this does seem too much to me. Especially on such a broad topic on which many research groups around the world have published. The subsection where the self-citing is most striking is on “Deep Learning”. The only deep learning model that is discussed is from the authors, whereas it is not the first paper on this topic. Why e.g. is not [3] discussed?

→ We absolutely agree with your comment. Therefore, we described more studies on deep learning-based lung cancer prognostication. Specifically, we described the following studies: 1) a model that could provide a prognostic score for the additional survival benefit from tyrosine kinase inhibitors and immune checkpoint inhibitors among stage IV lung cancers; 2) a model that could measure the solid portion volume in early-stage lung cancers, which showed high agreement with manual measurements by radiologists; 3) deep learning-based measurements of the solid portion volume, which showed higher prognostic discrimination than manual diameter measurements and traditional volumetric analysis. In addition, we presented an AI algorithm for the prediction of visceral pleural invasion that showed radiologist-level diagnostic performance.

Changes in the text:

Page 9, line 7-23

Deep learning-based prognostic CT features

Deep learning-based prognostication is also feasible in advanced-stage lung cancers. Deng et al. (35) recently proposed a deep learning model using pretherapy CT scans, which provided a

probability score to identify low-risk or high-risk patients receiving tyrosine kinase inhibitors and immune checkpoint inhibitors. The model was able to identify which patients would receive additional survival benefits beyond the median progression-free survival.

Deep learning is not limited to obtaining the cumulative survival probability; this method can be applied to capture the semantic features of lung cancer more broadly. Ahn et al. (41) reported that automatic measurements of the solid portions of lung cancer were comparable with manual measurements made by radiologists (intraclass correlation coefficient, 0.82-0.89) and showed good agreement with the invasive component size on microscopic examinations (intraclass correlation coefficient, 0.67). Kawaguchi et al. (42) showed that the solid component volume measured using deep learning had higher prognostic discrimination for recurrence or death (AUC, 0.752) than the solid component size (AUC, 0.722) or traditional three-dimensional volumetric analysis (AUC, 0.723). The presence of visceral pleural invasion was also estimated using preoperative CT scans (31). In a study by Choi et al. (31), the diagnostic performance of the deep learning model (AUC, 0.75) for visceral pleural invasion was on par with that of board-certified radiologists (AUC, 0.73-0.79).

Minor comments:

3. Line 68: “A recent study reported that 68 measuring multiple solid components, despite substantial inter-reader variability, did 69 not lead to better prognostication of stage IA lung adenocarcinomas (7)”.

The word ‘despite’ seems out of place here because it suggests a (seeming) contradiction, whereas “substantial inter-reader variability” can likely contribute to the feature not being a strong prognostic factor. So what is the contradiction?

→ We clarified the sentence as follows. Thank you for your comment.

Changes in the text:

Page 4, line 8-9

A recent study reported that measuring multiple solid components did not improve predictions of the prognosis of stage IA lung adenocarcinomas (10).

4. In “Tumor dimensions” no mention is made of gross-tumor value (primary plus metastases), although for stage III this is one of the more important prognostic factors [1]

→ In accordance with your comment, we described the gross tumor volume as follows.

Changes in the text:

Page 4, line 16-20

Gross tumor volume, encompassing both the volume of the primary tumor and metastases, is also an important prognostic factor (1). A recent meta-analysis on the prognostic factors for overall

survival in patients with stage III non-small cell lung cancer found that gross tumor volume was significant in 7 out of 9 multivariable analyses (1).

5. Line 102: “insignificant after adjustment for confounders (14, 15)”. Confounders are relevant when making inferences about the causal effect of one variable (or treatment) on another. When it comes to infering wheter a feature has “independent prognostic value”, the term confounder is irrelevant and also inaccurate as e.g. mediators may also reduce prognostic value of the variable under consideration but mediators are not confounders. It would be better to say e.g. “insignificant after including other prognostic factors”.

→ Thank you for your correction. We revised to follow your suggestion.

Changes in the text:

Page 5, line 21-23

Nevertheless, debate continues regarding the validity of GGO as a prognostic variable. Some studies have argued that its prognostic value was insignificant after including other prognostic factors (17, 18).

6. Line 129: “Pleural retraction” is a different feature than margin spiculation and has a different physiological background. It may improve readibility by starting a new subparagraph for this sentence to make the story clearer.

→ We followed your suggestion.

Changes in the text:

Page 7, line 1-9

Pleural retraction or tag

Pleural retraction or a pleural tag on CT is a predictor of visceral pleural invasion, which is an established prognostic factor and a T2 descriptor in lung cancer. Nevertheless, CT definitions of pleural tag or retraction are highly variable among radiologists and often yield false-positive diagnoses for pathological visceral pleural invasion. Kim et al. (23) investigated combinations of CT findings, including tumor contact with the pleura, pleural retraction, and tags. None of the combinations were associated with recurrence-free survival in clinical T1N0 lung cancers. Therefore, it is uncertain whether these CT features can be reliably identified and whether these features are genuinely indicative of patients’ survival.

7. Line 142: “Central tumor location is an indicator of occult nodal metasasis”. The word ‘indicator’ might be too strong because it is not a one on one relationship, perhaps ‘is associated with” is a more correct phrasing.

→ We followed your suggestion.

Changes in the text:

Page 7, line 11-12

Central tumor location is associated with occult nodal metastasis and mediastinal nodal disease in radiologically node-negative, early-stage lung cancers (24).

References

1: van Laar M, van Amsterdam WAC, van Lindert ASR, de Jong PA, Verhoeff JJC. Prognostic factors for overall survival of stage III non-small cell lung cancer patients on computed tomography: A systematic review and meta-analysis. Radiother Oncol. 2020 Oct;151:152-175. doi: 10.1016/j.radonc.2020.07.030. Epub 2020 Jul 22. PMID: 32710990.

2: Wu L, Lou X, Kong N, Xu M, Gao C. Can quantitative peritumoral CT radiomics features predict the prognosis of patients with non-small cell lung cancer? A systematic review. Eur Radiol. 2022 Oct 29. doi: 10.1007/s00330-022-09174-8. Epub ahead of print. PMID: 36307554.

3: Hosny A, Parmar C, Coroller TP, Grossmann P, Zeleznik R, Kumar A, Bussink J, Gillies RJ, Mak RH, Aerts HJWL. Deep learning for lung cancer prognostication: A retrospective multi-cohort radiomics study. PLoS Med. 2018 Nov 30;15(11):e1002711. doi: 10.1371/journal.pmed.1002711. PMID: 30500819; PMCID:PMC6269088.

Reviewer C

This is a very concise review on CT features for prognosication of NSCLC. It covers various topics from morphologiy to radiomics.There are many self citations. But, that would be acceptable.

→ We appreciate that you took the time to review this manuscript and your comments on our study.

Review Comments (Round 2)

Reviewer A

The authors have addressed my remarks and the manuscript has improved. I found one typo: page 9, line 273 has the word "pretherapy" which should be split.

→ Thank you very much for your review. We followed your comment and revised the word from “pretherapy” to “pre-therapy”.