Peer Review File

Article information: https://dx.doi.org/10.21037/tcr-21-702

Reviewer A

In this manuscript, the authors conducted a retrospective study. The tried to incorporate radiomic features of CT and clinical factors to build a nomogram. The diagnostic accuracy of the nomogram in predicting brain metastasis of stage IV lung cancer patients was compared to that of the brain MRI. They reported that the area under curve of the nomogram was 0.911.

Comment 1:

Novelty of this study:

In recent years, the nomograms built with Rad-score in predicting occurrence of cancer or survival of patients have been reported by several groups.

Reply 1: Thanks. The purpose of this study is to build a nomogram by incorporating radiomic signature (Rad-Score) and clinical risk factors to predict the occurrence of occult brain metastasis. This is a clinical-oriented study with widely used techniques in previous studies.

Comment 2:

Introduction:

The occult brain metastasis was diagnosed by CT in this study. Nowadays, the brain MRI is recommended to scan brain metastasis in most guidelines. It seems too inappropriate to say a brain metastasis missed by CT is “occult”.

Reply 2: Thanks for your valuable comments. In this study, “occult” refers to the case where brain metastasis is not found in the CT report but is found in MRI. Just like occult lymph node metastasis means that lymph node metastasis is not reported on the chest CT, but the presence of lymph node metastasis is confirmed by surgery [1-2]. Given that, we have added the description in context, please see line 69-71, page 3.

[1] Zhong Y, Yuan M, Zhang T, Zhang YD, Li H, Yu TF. Radiomics Approach to Prediction of Occult Mediastinal Lymph Node Metastasis of Lung Adenocarcinoma. AJR Am J Roentgenol. 2018 Jul;211(1):109-113.

[2] Jiang L, Jiang S, Lin Y, Yang H, Xie Z, Lin Y, Long H. Nomogram to Predict Occult N2 Lymph Nodes Metastases in Patients With Squamous Nonsmall Cell Lung Cancer. Medicine (Baltimore). 2015 Nov;94(46):e2054.

Comment 3:

Patients and Methods:

(1) This first inclusion criteria were no brain metastasis observed in pretreatment baseline chest CT. CT of chest protocol is not suitable to detect brain meta.

(2) The patients underwent “unenhanced” cheat CT for tumor staging. “Unenhanced” chest CT is not adequate for accurate staging of lung cancer patients.
(3) Was the brain was scanned during unenhanced chest CT? The description for CT images acquisition should be in more detail.
(4) Because MRI machine is available in most cancer centers. The definition of occult brain metastasis should be those missed by MRI. The clinical impact of this study is doubtful because a brain MRI scan is already recommended for stage IV lung patients. Moreover, unenhanced chest CT is not adequate to determine TN status, the accuracy of tumor staging of the study patients is also doubtful.
Reply 3: Thanks for your valuable comments.
(1) Surely, CT of chest protocol is not suitable to diagnose BM. In this study, we didn't make it clear, so we revised the inclusion criteria, ① clinical diagnosed with stage IV lung cancer; ② no BM observed in pretreatment baseline cranial CT images; ③ have cranial MRI and baseline chest CT. CT images is only used to extract radiomic features and build signature and nomogram, and cranial MRI is used to determine the status of BM for stage IV LADC. Please see line119-124, page 7.
(2) In this study, staging of lung cancer patients is based on whether the patient has distant metastasis (including brain metastasis). We have revised the description, please see line 118-119, page 6.
(3) In this study, we extracted quantitative imaging features from primary lung tumor to decode its heterogeneity. Therefore, no brain was scanned during chest CT. In our institution, the chest and brain are scanned separately using different imaging protocols. We have added more details about CT imaging in context, please see line 133-140, page 7.
(4) Although pathology is considered as “gold standard” for diagnosing BM, it is generally diagnosed by cranial MRI due to its special location rather than biopsy. Although MRI is already recommended for stage IV lung patients, however, it was unavailable for any patient with heart pacemaker, claustrophobia, metal implant or some absolute contraindications, or patients with low KPS due to long scanning time. We have stated and revised this reason in context, please see line 78-81, page 5. Besides, all patients with lung cancer can receive thoracic CT examination, but not all patients can do craniocerebral MRI. Therefore, The purpose of our study is to evaluate the probability of BM through the radiomic features derived from primary lesion of chest CT, further to determine the high-risk groups prone to BM. The stage of tumor depends on clinical diagnosis not only chest CT, and once there is distant metastasis

Reviewer B

This is an interesting study using CT radiomic features to predict occult brain metastasis (BM) in patients with stage IV lung adenocarcinoma. Consecutive patients were separated into a training and validation cohort and 725 radiomic features extracted from the lung tumor on CT. Variability of the features was assessed through

simulated perturbations of the contours. LASSO was used to select the optimal radiomics features to predict occult BM in the training cohort. A radiomics score and subsequent nomogram incorporating clinical features was then developed and evaluated. Overall, the results are quite promising. However, I have several concerns regarding the unmet need of this paper, patient population used, and overall methodology details that need to be clarified.

Comment 1: Introduction: "Notably, the presence rate of BM for advanced stage patients with NSCLC were higher than earlier stages." This statement is unclear. If BM is present, then the patient is automatically advanced stage. Therefore, the intent of this sentence is unclear.
Reply 1: Thanks for your valuable comment. The intend of this sentence is the presence rate of BM for advanced stage patients with NSCLC who have bone metastasis, liver metastasis, or/and adrenal metastasis (stage IV patients with no BM) were higher than earlier stages. We have rephrased this statement, please see line 69-71, page4.

Comment 2: The unmet need that this paper addresses are unclear. The introduction needs a clearly described and well-cited description of the unmet need and that it is of high importance. For example, the authors state that the "early detection of occult BM and delivering a tailor therapy can significantly improve prognosis of stage IV LADC." These statements are insufficiently specific and there are no citations backing up these claims. The introduction needs a carefully cited argument stating:
(a) What the current diagnostic methods for stage IV lung cancer are (and specifically for identifying BM) and how well they perform?
(b) The specific clinical impact/decision that finding radiomics associations with BM would support. For example, what tailored treatments or individual management would be available for patients. There needs to be a cited argument for the impact on clinical care will this have.
(c) The percentage of patients impacted by (b) above. The authors clearly state the incidence of NSCLC patients with BM at first diagnosis, but how many patients have occult BM? In general, what is the cited negative impact on this patient population.
Reply 2: Thanks for your valuable comments.
(a) Guidelines for NSCLC raised by NCCN recommend MRI for the detection of single or multiple intracranial BM. However, although cranial MRI was considered as gold standard for diagnosing occult BM, it was unavailable for any patient with heart pacemaker, claustrophobia, metal implant, some absolute contraindications, or low KPS. We have added some necessary statements and cited references, please see line 78-81, page 5.
(b) We have revised our statement on tailor treatment of BM and cited necessary references to support our arguments. Please see line 73-75, page 4.
(c) To our experience, the presence rate of occult BM is approximate 40% in our institution. However, there is no published paper or official statement have reported

the rate of occult BM. Therefore, we have not stated the incidence of occult BM in context. But we have listed the numbers of occult BM in training and validation cohort in Table 1.

Comment 3: Methods – Patient Population: Overall the patients included in this study is unclear and more specific details are needed. In general, the definition of an occult BM needs to be clearly defined.
(a) Is cranial MRI standard of care for all patients at your institution with stage IV LADC? If not, then requiring a patient to have a cranial MRI as part of the inclusion criteria is going to lead to bias in the results.
(b) Similarly, it was stated that no BM was observed in the pre-treatment baseline chest CT. However, a chest CT would not include the skull, therefore how can you determine if a BM was present or not?
(c) Overall, additional details are needed on the patient demographics. For example, what was the time between initial diagnosis of the lung tumor and identification of the brain metastases?
Reply 3: Thanks for your valuable comments.
We have added description about the definition of occult BM. Please see line 66-68, page 3.
(a) Yes, cranial MRI is standard of care for all patients at your institution with stage IV LADC. However, patients will be prohibited from MRI examination because of some absolute contraindication descripted in line 78-81, page 5. This study will eventually benefit these patients who cannot receive MRI scanning.
(b) Sorry, we have made a mistake. The correct sentence should be “no BM observed in pretreatment baseline cranial CT images”. In our institution, the chest and brain are scanned separately using different imaging protocols. Therefore, no brain was scanned during chest CT. We have revised this error. Please see line 120-121, page 7.
(c) The occult BMs of stage IV LADC were detected at first diagnosis. Therefore, the time interval between first diagnosis and identification of BM is usually within a week for patients with occult BMs. For patients without BMs at initial diagnosis, the time interval between first diagnosis and develop BM are varied. Since this study was retrospective, we have included all available characteristics in Table 1.

Comment 4: Methods – Feature Extraction: All feature extraction parameters and / or pre-processing techniques need to be described for reproducibility.
Reply 4: Thanks for your valuable comment. In this study, we have calculated the feature reproducibility against segmentation variability.

Comment 5: Methods – Segmentation: It is unclear how the interobserver variance was assessed.
(a) The choice of these parameters needs to be justified. For example, are these producing realistic contours that would be expected by another radiologist? A figure may be beneficial to demonstrate this.
(b) Why wasn’t a second radiologists asked to manually contour a subset of images?

(c) I understand this was done for a subset of 30 patients. How was this number determined and how were these patients selected?

Reply 5: Thanks for your valuable comments.

(a) and (b): Figure R1 is the workflow of multiple segmentation test. We have added a flow diagram to clarify your concerns. Thanks for your suggestion.

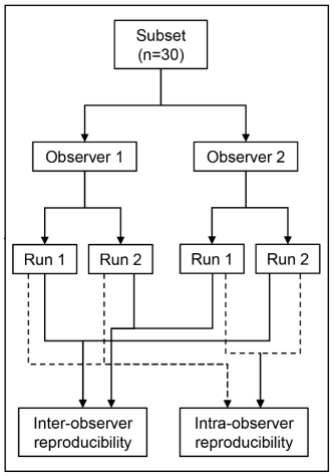

Figure R1. Multiple segmentation test to evaluate reproducibility of radiomic features.

(c) As a rule of thumb, to examine the reproducibility of radiomic features and reduce the false discovery rate, datasets consisting of 10–15 patients per feature evaluated have been recommended [1-2]. Our previous study utilized 26 patients to calculate the reproducibility and reduce redundant of radiomic features against multiple semi-automatic segmentation methods [3]. Therefore, 30 patients were enrolled in subset.

[1] Chalkidou A, O'Doherty MJ, Marsden PK. False discovery rates in PET and CT studies with texture features: a systematic review. PLoS One 2015;10: e0124165.

[2] Yip SS, Aerts H. Applications and limitations of radiomics. Phys Med Biol 2016;61: R150-66.

[3] Qiu Q, Duan J, Duan Z, et al. Reproducibility and non-redundancy of radiomic features extracted from arterial phase CT scans in hepatocellular carcinoma patients: impact of tumor segmentation variability. Quant Imaging Med Surg 2019;9(3):453-464.

Comment 6: Methods: Why was an ICC < 0.75 use to determine robust features? This choice needs to be justified, as in most radiomic studies use a higher cut-off value.

Reply 6: Thanks for your valuable comment. Sorry to made this error. A feature with ICC≥0.75 was considered as robust and reproducible feature according to published papers. We have corrected this error, please see line 171, page 9.

Comment 7: Methods: It is unclear what analysis was performed on the training cohort and what on the validation cohort. The "Radiomics Signature and Nomogram

Construction" section does not specify what dataset each step was performed on. It is unclear if there was information leakage between the training and validation cohorts. A flow diagram clearly representing each step within your methodology for each cohort would help clarify this.
Reply 7: Thanks for your valuable comments. We have added a flow diagram in Supplementary Figure S1. That is quite a nice suggestion. Thanks again.

Comment 8: Table 1: The *P* values listed in the text should be included directly in the table for each factor.
Reply 8: Thanks for your valuable comment. We have added a new column in Table 1 with title "*P* value".

Comment 9: Results – Feature Robustness: The authors state that "549 out of 725 extracted features have ICC ≥ 0.75, indicating that high reproducibility among intra- and inter-observer multiple segmentation test." However, intra-observer variability was not assessed since simulations were performed on the segmentations to mimic variations.
Reply 9: Thanks for your valuable comments. In section of feature robustness, intra-observer multiple segmentation test was performed. Two segmentations were finished by one radiologist with time interval 5 days. Since lung tumor have distinct boundary, two delineations (run 1 and run 2 of observer 1, or 2) are almost same. For intra-observer reproducibility, 703 out of 725 (97%) features have ICC≥0.75. Both ICC derived from intra- and inter-observer are both higher than 0.75, the feature can be considered as robust. As a result, 549 out of 725 extracted features have ICC ≥ 0.75, indicating that high reproducibility among intra- and inter-observer multiple segmentation test.

Comment 10: Results – LASSO: Additional details are needed on how LASSO was performed. Was this a cross-validated LASSO (cv.glmnet)? All parameters need to be described for reproducibility (e.g., were features standardized).
Reply 10: Thanks for your comment. We have added necessary parameters to ensure reproducibility. Please see line 173-176, page 9.

Comment 11: Results: What dataset was the univariate analysis of the clinical features performed on? This is related to comment 7 above regarding clarity of methods.
Reply 11: Thanks for your comment. We have added the statement "The univariate analysis was performed on training cohort." to clarify the concern. Please see line179-180, page 9.

Comment 12: There are several grammatical errors throughout. Please proofread carefully.
Reply 12: Thanks for your comment. We have polished our language by AME Editing Service.