Peer Review File

Article information: https://dx.doi.org/10.21037/jtd-23-1578

Reviewer A

The authors presented a comparison of different machine learning for Radiomics classification. The paper is not ready for publication.

1)An assertion "Extreme gradient boosting is the best algorithm for radiomics modeling" is questionable. The authors did try experiment with a small dataset from just one center. Multiple datasets are required to validate this claim.

Reply 1: We do agree that the extreme gradient boosting (XGB) may not be the best algorithm for radiomic modeling in general. However, in this study, we focused on predicting occult lymph node metastasis (ONM) in clinical stage IA lung adenocarcinoma patients. We compared XGB with other machine learning methods using various metrics, such as accuracy, sensitivity, specificity, F1-score, and AUC, and found that XGB outperformed the others in our data set. Therefore, we concluded that XGB is the best algorithm to build radiomics model for predicting ONM in patients with clinical stage IA lung adenocarcinoma. We apologized for the inaccurate statement "Extreme gradient boosting is the best algorithm for radiomics modeling, and a combined model of radiomics and clinical-radiological features can better predict occult lymph node metastasis in stage IA lung adenocarcinoma." in the manuscript for peer-review. We have reorganized the sentence in the revised manuscript as "Extreme gradient boosting is the best algorithm to build radiomics model for predicting occult lymph node metastasis (ONM) in patients with clinical stage IA lung adenocarcinoma, yielding an Area under curve (AUC) of 0.917, and the combined model incorporating radiomics features and clinical-radiological features can better predict ONM with a superior AUC of 0.933". (see Page 5, line 57)

We strongly agree that multiple data sets are needed to validate its performance and generalizability. We have stated that the relatively small sample size due to the low ONM rate in clinical stage IA lung adenocarcinoma, and multicenter studies may be adopted for validation in clinical applications in the limitation section (see Page 16, line 275-276). We are already beginning to collect data from multiple centers in the hope of building a clinically applicable universal model.

2) No novelty in this paper.

Reply 2: We disagree with the notion that our paper lacks novel viewpoints. While there have been studies on predicting lymph node metastasis in lung cancer using radiomics-based machine learning models, the diagnostic performance varies across different research models (Table 1). Our paper firstly distinguishes itself by conducting a comprehensive comparison of seven machine learning algorithms and assessing their performance using various metrics. We believe that this in-depth analysis provides valuable insights into the strengths and limitations of each algorithm, aiding in the identification of the optimal machine learning algorithm for radiomics-based prediction of ONM in clinical stage IA lung adenocarcinoma patients.

Reference:
1. Zhong Y, Yuan M, Zhang T, Zhang Y-D, Li H, Yu T-F. Radiomics Approach to Prediction of Occult Mediastinal Lymph Node Metastasis of Lung Adenocarcinoma. American Journal of Roentgenology. 2018;211(1):109-13.
2. Cong M, Feng H, Ren JL, Xu Q, Cong L, Hou Z, et al. Development of a predictive radiomics model for lymph node metastases in pre-surgical CT-based stage IA non-small cell lung cancer. Lung Cancer. 2020;139:73-9.
3. Zheng X, Shao J, Zhou L, Wang L, Ge Y, Wang G, et al. A Comprehensive Nomogram Combining CT Imaging with Clinical Features for Prediction of Lymph Node Metastasis in Stage I-IIIB Non-small Cell Lung Cancer. Ther Innov Regul Sci. 2022;56(1):155-67.

3)Radiological features: why averaging the features? What about the experts agree on a certain value for each feature by discussion?

Reply 3: Thank you for review. In our methodology, we chose to calculate the average values of the quantitative characteristic (diameter) measured by two radiologists as a strategy to address individual variations in measurements. Averaging the quantitative feature was driven by our intention to create a more robust and representative dataset, thereby minimizing the impact of inter-observer variability. Expert group discussion was adopted when there were discrepancies in interpretation of morphological features between two radiologists. Morphological features included nodule consistency (pure ground glass nodule [pGGN], part solid nodule [PSN], solid nodule [SN]), bubble-like lucency, bronchiectasis, deep lobulation, emphysema, necrosis, pleural indentation, shape, lobar location, and tumor location (central [the inner third] or peripheral [the outer two-thirds of the lung fields]). All these features are qualitative and subjective, and the group discussion can reach agreement through in-depth analysis of the controversial features. The detail of radiological feature interpretation was in Page 9-10, line 142-151.

Table 1: The diagnostic efficiency of different models in predicting lymph node metastasis of lung cancer

Machine learning algorithms	AUC	Reference
Support vector machine	0.972	1
Random forest	0.860	2
Logistic regression	0.700	3

4)Introduction is small and no clear research question is provided nor a background about the research in this area

Reply4: Thank you for review. We revised the introduction according to your comments. In the revised manuscript, the introduction consists of four parts. In the first paragraph, we introduced the clinical significance of preoperative diagnosis of ONM. In the second paragraph, we reviewed the current non-invasive methods for diagnosing ONM, expounded the advantages of radiomics and analyzed the reason why radiomics model for ONM diagnosis are difficult to apply widely in clinical practice. In the third paragraph, we outlined the utilization of machine learning methods for constructing radiomics models, emphasizing the existing variability in model selection across different studies and its impact on diagnostic consistency. And we pointed that currently no research has studied which machine learning algorithm is more suitable for establishing ONM diagnostic models. In the last paragraph, we state the purpose and clinical significance of this study (see Page 5-7, line 60-105).

5)Radiomics features were reduced to 6, but no comparison of the ML on different reduction of features was performed.

Reply 5: Thank you for your carefully review. At the very beginning, we would like to provide additional clarity on our methodology. Our study employed a two-step feature selection process involving the minimum redundancy maximum correlation (mRMR) method followed by the Least Absolute Shrinkage and Selection Operator (LASSO) method to select radiomics features. We did not use multiple machine learning methods in the feature screening, because mRMR and LASSO had well-established effectiveness in feature selection which evidenced by studies in the literature (1-3). Different feature selection methods may affect the final retained features, but our main purpose of this study is to explore which ML method is the best method to build radiomics model for ONM diagnosing in patients with clinical stage IA lung adenocarcinomas. We adopted seven machine learning methods in model building and compared the clinical diagnostic capabilities of the models they built. Finally, we found that XGB is the best algorithm to build radiomics model for predicting ONM in patients with clinical stage IA lung adenocarcinoma.

References:

1. Wu Z, Wang H, Zheng Y, Fei H, Dong C, Wang Z, et al. Lumbar MR-based radiomics nomogram for detecting minimal residual disease in patients with multiple myeloma. European Radiology. 2023;33(8):5594-605.
2. Hou J, Li H, Zeng B, Pang P, Ai Z, Li F, et al. MRI-based radiomics nomogram for predicting temporal lobe injury after radiotherapy in nasopharyngeal carcinoma. European Radiology. 2022;32(2):1106-14.
3. Bi S, Li J, Wang T, Man F, Zhang P, Hou F, et al. Multi-parametric MRI-based radiomics signature for preoperative prediction of Ki-67 proliferation status in sinonasal malignancies: a two-centre study. European Radiology. 2022;32(10):6933-42.

6)What are the six features that were used for ML?

Reply 6: We appreciate your careful review of our manuscript. Regarding your inquiry about the six features used in our machine learning approach, we would like to direct your attention to Table S3 in the Supplemental material, where a detailed listing of these features can be found.

Reviewer B

Great job, well-written article! Nothing to add.

Reply 1: Thank you for your review and recognition of our research.

Reviewer C

• Background Explanation (29-32):

Suggestion: It might enhance the context to consider adding a sentence in the background, briefly explaining the clinical impact of accurately predicting occult lymph nodes. For instance, you could mention how precise predictions may significantly influence treatment decisions and positively impact patient outcomes.
Reply 1: We appreciate your feedback and agree that providing additional context in this regard would enhance the overall understanding of our research. According to your suggestion, we have revised the manuscript to include the following sentence in the background section: "Accurate prediction of occult lymph node metastasis (ONM) is an important basis for determining whether lymph node dissection is necessary in clinical stage IA lung adenocarcinoma patients." (see Page 3, line 27-29)

• Avoid Repetition (29-30):
Suggestion: Could we possibly avoid the repetition of the terms 'radiomics modeling' and 'radiomics model' in this section for a smoother flow?
Reply 2: Thank you for your feedback. We have addressed the repetition of 'radiomics modeling' and 'radiomics model' in the section. The revised sentence is "The aim of this study is to determine the best machine learning algorithm for radiomics modeling and to compare the performances of the radiomics model, the clinical-radiological model and the combined model incorporating both radiomics features and clinical-radiological features in preoperatively predicting ONM in clinical stage IA lung adenocarcinoma patients." (see Page 3, line 29-32)

• Clinical Relevance Emphasis (Entire Abstract):
Suggestion: Throughout the abstract, consider emphasizing the clinical relevance of the study's conclusion. Highlight how the superior performance of the combined model in predicting occult lymph node metastasis (ONM) could notably influence and improve clinical decision-making.
Reply 3: Appreciate your insightful feedback. We have incorporated your suggestion into the abstract, underscoring the clinical relevance of our study's conclusion. Specifically, I included the following statement in the conclusion on Page 4, Line 47-50: "The superior performance of the combined model based on extreme gradient boosting algorithm in predicting ONM in patients with clinical stage IA lung adenocarcinoma might aid clinicians in deciding whether to conduct mediastinal lymph node dissection and contribute to improve patient's outcomes."

• Highlight Box Enhancement:
Suggestion: For the highlight box, it may be beneficial to consider adding some specific key points that would provide further clarity and enhance understanding.
Reply 4: Thank you for your feedback. We appreciate your suggestion regarding the highlight box for our paper. In response, we've included the performance of Extreme gradient boosting model and combined model in the highlight box (see Page 5, line 57).

• Word Choice Adjustment (57):
Suggestion: To enhance clarity and precision, consider adjusting 'deviate' to 'differ' in this context.
Reply 5: Thanks to your suggestion, we have adjusted "deviate" to "differ" in the manuscript (see Page 5, line 63).

• Generalization in Standard Treatments (60-61):
Suggestion: For generalization, you might consider omitting 'The standard treatment' and instead using 'Standard treatments.'
Reply 6: Thanks for your feedback, we have modified it. The revised sentence is "To ensure sufficient resection of lymph node lesions, systematic lymph node dissection is still recommended as standard method for early-stage lung cancer patients." (see Page 6, line 66-67)

• Specify Radiomics Modeling Goals (80):
Suggestion: If applicable, it could be beneficial to specify the goals or specific aspects of radiomics modeling that the study aimed to optimize, such as feature selection or model training parameters.
Reply 7: Your insights have been invaluable in refining the focus of the study. We would like to clarify that the aim of our research is compare different machine learning algorithms to build radiomics model to predict ONM in stage IA adenocarcinoma. Although there have been studies based on radiomics model to predict lymph node metastasis in lung cancer, the diagnostic efficacy of these studies varies widely, as shown in Reply to reviewer A-2. This is mainly because the data types targeted by the different machine learning models are not consistent, such as logistic regression models are more suitable for linear data, while extreme gradient boosting models are more suitable for nonlinear data[1,2]. Therefore, it is necessary to verify which model is more appropriate for studying a specific clinical problem.

References:
1. Subasi A. Chapter 3 - Machine learning techniques. In: Subasi A, editor. Practical Machine Learning for Data Analysis Using Python: Academic Press; 2020. p. 91-202.
2. Cervellera C, Macciò D. Gradient Boosting with Extreme Learning Machines for the Optimization of Nonlinear Functionals. In: Paolucci M, Sciomachen A, Uberti P, editors. Advances in Optimization and Decision Science for Society, Services and Enterprises: ODS, Genoa, Italy, September 4-7, 2019. Cham: Springer International Publishing; 2019. p. 69-79.

• Detailed Comparison Metrics (81):
Suggestion: Please consider providing a brief mention of the specific metrics or criteria used to compare the performance of the machine learning algorithm and the clinical model. Metrics like the area under the curve, sensitivity, and specificity could be briefly highlighted.
Reply 8: Thank you for your advice. We have revised the manuscript to include the following sentence: "Evaluation metrics, including AUC, accuracy, precision, sensitivity, specificity, and F1 score, were specifically employed to measure and compare the performance of these models." (see Page 7, line 101-105)

• Emphasize Potential Significance (Introduction):
Suggestion: Throughout the abstract, consider elaborating on the potential significance of the novel insights and valuable strategies.
Reply 9: Appreciate your insightful feedback. We have revised the manuscript to include the

following sentence: “The insights gained will provide valuable strategies for preoperative ONM prediction in clinical stage IA lung adenocarcinoma, contributing to improved clinical decision-making and patient prognosis.” (see Page 7, line 103-105)

• Acknowledgment of Limitations (Introduction):
Suggestion: Consider adding a brief acknowledgment of any limitations or challenges encountered during the study.
Reply 10: Thank you for your advice. We have incorporated the following acknowledgment of the study's limitations into the manuscript: “Researches focus on ONM diagnosis have achieve as high as 0.972 of area under curve (AUC) by using radiomics-based diagnostic model. However, Zhang et al.'s research reported that the AUC of radiomics model in diagnosing ONM was only 0.813. The inconsistence in diagnostic efficacy hinders the clinical application of radiomics models.” (see Page 6, line 81-84)

• Numerical Clarity (101-102):
Suggestion: Consider changing the last sentence so as to clarify the final patient distribution.
Example: "Ultimately, this study included 258 patients (129 LN-positive and 129 LN-negative), randomly divided into training and test cohorts in a 7:3 ratio."
Reply 11: Thank you for your valuable feedback. To provide a clearer understanding of the patient distribution, we have revised the sentence as follows: “Ultimately, this study included a total of 258 patients, with 182 patients (91 LN-positive and 91 LN-negative) assigned to the training set and the remaining 76 patients (38 LN-positive and 38 LN-negative) forming the test set. The random division followed a 7:3 ratio, ensuring a representative distribution across both sets.” (see Page 8, line 121-124)

• List Clarity (107-108):
Suggestion: Consider formatting the list of tumor grades for improved clarity.
Reply 12: Thanks for your suggestion, we have displayed the list of tumor grades in the Table S2 in supplemental materials.

• Wording Adjustment (Lines 113-117):
Suggestion: Restructure sentences for clarity, e.g., "HRCT images were acquired using 8- (LightSpeed Ultra, GE Medical Systems), 16- (ProSpeed or Discovery ST, GE Medical Systems), or 64- (LightSpeed VCT, GE Medical Systems) spiral CT scanners."
Reply 13: Thank you for your feedback. We have restructured the sentence to improve readability: “We acquired HRCT images with 8-, 16-, or 64-spiral CT scanners (LightSpeed Ultra, ProSpeed, Discovery ST, or LightSpeed VCT; GE Medical Systems).” (see Page 9, line 136-137)

• Experience Clarity (Lines 120-121):
Suggestion: Clarify the experience of the radiologists for enhanced understanding.
Example: "with over 10 years of experience in chest CT interpretation."
Reply 14: Thank you for your feedback. We have modified it in the manuscript. The revised sentence is “The radiological features of the primary tumor were assessed by two experienced

thoracic radiologists (L.Z. and M.L.) both with over 10 years of experience in chest CT interpretation.” (see Page 9, line 142-143)

• Link Clarification (Line 144):
Suggestion: Instead of providing the URL, mention that the formula for calculating radiomics signatures can be found on the official documentation website or add a hyperlink.
Reply 15: Thank you for your valuable feedback. The suggested modifications have been implemented in the manuscript (see Page 10, line 164).

• Machine Learning Algorithms (Lines 150-151):
Specify the acronym expansion for clarity.
Example: Expand acronyms for machine learning algorithms, e.g., "Bayes, LASSO, LR, DT, RF, SVM, and XGB."
Reply 16: Thank you for your suggestion. We would like to clarify the full term “Bayes, LASSO, LR, DT, RF, SVM, and XGB” is introduced for the first time in Page 7, line 85-87.