Can CT radiomics differentiate benign from malignant N2 adenopathy in non-small cell lung cancer

Robert J. Cerfolio[1], William H. Moore[2]

[1]Department of Cardiothoracic Surgery, New York University Langone Health, New York, NY, USA; [2]Department of Radiology, NYU Langone Health, New York, NY, USA

Correspondence to: Robert J. Cerfolio, MD, MBA, FACS, FCCP. Executive Vice President of New York University (NYU) Langone Health; Vice Dean, Professor of Cardiothoracic Surgery NYU; Chief of Thoracic Surgery; Director of the Lung Cancer Service Line, New York, NY, USA. Email: Robert.Cerfolio@nyulangone.org; William H. Moore, MD. Associate Professor of Radiology, Section Chief, Thoracic Imaging, Associate Chair of Clinical Informatics, Department of Radiology, NYU Langone Health, 660 1st Avenue, 3rd Floor, New York, NY 10016, USA. Email: william.moore@nyulangone.org.

Comment on: Yang M, She Y, Deng J, *et al*. CT-based radiomics signature for the stratification of N2 disease risk in clinical stage I lung adenocarcinoma. Transl Lung Cancer Res 2019;8:876-85.

We congratulate Dr. Yang and the other authors for their interesting paper on stratifying N2 disease using CT based radiomics signature (1). The team has taken on an important challenge of differentiating benign from malignant lymph nodes based on radiomics markers rather than depending on size alone. We have several questions as well as a few comments, which focus on two primary areas. The first is the clinical applicability of this paper and second on its methods.

The authors did not use integrated 18F-fluorodeoxyglucose (FDG) positron emission tomography (PET) computed tomography (CT) scan in this study. We are quite surprised by this for a study on lung cancer and the accuracy of clinically staging of N2 disease in 2014 and 2015. A large meta-analysis performed in 2012 showed that combined integrated PET/CT have higher specificity with summary receiver operating characteristic (SROC) of 0.8067 the CT scan alone (2) This omission potential flaws the entire study since we know that integrative Pet CT is better at detecting N2 disease than CT scan alone. The decision to not include PET/CT as part of this study requires a robust explanation in the discussion. If some and not all patients underwent PET/CT can they tell us the results for those that did?

Second, although this article is going into a radiologic journal, surgeons play a role in this study. Notably missing from the article is the type of lymph node dissection performed. I have had the great honor of being an invited visiting professor at The Shanghai Pulmonary Hospital and I know some of the very many outstanding surgeons there. From both watching and operating there, I know that some practice selected lymph node sampling, others perform complete thoracic lymphadenectomy and yet a third group routinely perform a complete lymph node sampling. The variation in practice results in significant bias. If the gold standard by which we determine pathologic N2 disease is different and if all N2 disease are not completely removed in all patients, the entire series is flawed. The gold standard of pathologic N2 is variable. Finally, there is a difference between single station N2 disease and multi-station especially on the left side (5 and 6 stations). Can the authors expound on this as well?

Next, we have several questions about the methodology of the artificial intelligence (AI) itself. We appreciate the use of an open source software package, and we believe this may be the future of AI. However, there are issues with the methods as to how the AI system was implemented in this study. The group started with 1,078 radiomic markers and picked the top performing 100 markers. This is reasonable. They then used a least absolute shrinkage and selection operator (LASSO) regression mode to pick the best features. However, the authors then decided to use the top five performing features. Many articles related to

radiomic and texture features pick fewer markers because of the potential for over-fitting. The reason for choosing five features is not clearly explained in the article and should be.

The area under the curve (AUC) presented in this article is 0.81/0.82 and is not dissimilar to the performance of PET/CT prior to 2012. The authors then decided to add additional clinical parameters to their analysis of N2 lymph nodes. While we agree that this is an important aspect of the care of patients with lung cancer, the manner in which it was performed in this study remains unclear. AI systems promise the ability to integrate data in ways that we as human cannot. However, in these types of studies the rationale for including data requires clear explanation. Finally, since this a single institution study the potential for over-fitting of the data is real. Can the authors explain what data was omitted and which data was included to better enhance applicability at other institutions?

Acknowledgments

Footnote

Provenance and Peer Review: This article was commissioned by the editorial office, *Translational Lung Cancer Research*. The article did not undergo external peer review.

Conflicts of Interest: Both authors have completed the ICMJE uniform disclosure form (available at http://

dx.doi.org/10.21037/tlcr-20-477). WHM reports other from Merck, grants from Galil Medical, other from BTG, outside the submitted work; in addition, WHM has a patent for Scannerside issued outside of the submitted work. The other author has no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

1. Yang M, She Y, Deng J, et al. CT-based radiomics signature for the stratification of N2 disease risk in clinical stage I lung adenocarcinoma. Transl Lung Cancer Res 2019;8:876-85.
2. Zhao L, He ZY, Zhong XN, et al. (18)FDG-PET/CT for detection of mediastinal nodal metastasis in non-small cell lung cancer: a meta-analysis. Surg Oncol 2012;21:230-6.