Peer Review File
Article information: https://dx.doi.org/10.21037/tlcr-24-168

Reviewer A

The authors utilised microscopic Raman imaging coupled with convolutional neural network in evaluating for tumour cell invasion to distinguish minimally invasive adenocarcinoma and invasive adenocarcinoma in lung adenocarcinoma. While the concept is interesting, there are multiple major issues with this manuscript:

First, the Raman technique utilised samples that have already undergone pathology histologic processing with formalin fixation/paraffin embedded. It is unclear if Raman technique was performed on the same tissue block that was used from the intraoperative frozen analysis (i.e. the frozen remnant processed for pathology) or a different tissue block. Because if latter, it is just a sampling problem when one cannot diagnose minimally invasive adenocarcinoma on frozen, and it does not mean Raman was truly superior to intraoperative frozen section evaluation.

Reply 1: Thank you for your detailed review and valuable suggestions. In this study, samples were sourced from clinical settings, with each clinical sample undergoing meticulous documentation of intraoperative rapid pathology analysis and postoperative histopathological gold standard results by the pathology department. Selected samples were subjected to formalin fixation and paraffin embedding. To mitigate potential interference from dyes used in rapid pathology slides with Raman spectroscopy analysis, the serial section method was utilized for Hematoxylin and Eosin (H&E) staining, immunohistochemistry, and preserving unstained slides for Raman spectroscopy. This approach aims to maintain informational consistency across adjacent sections and reduce any sample selection bias that could affect the outcomes. The area for Raman spectroscopy collection was referenced against the postoperative histopathological standard images (H&E staining and immunohistochemistry), with all sections taken at the same coordinates. A schematic of this will be presented in the supplementary files. Given the lengthy preparation time of the postoperative histopathological standard, which does not meet the needs for intraoperative rapid diagnosis, we aim to achieve near-gold-standard diagnostic results intraoperatively with Raman technology, demonstrating its advantages over traditional frozen section methods.

Changes in the text: We have modified our text as advised (see Page 8, line196) or supplementary files will include detailed additions to this section.

Second, even if the identical tissue block (the frozen remnant) was evaluated, given the reported protocol with the Raman imaging only analysing three 200x200um2 regions, it

appears that this has to be chosen by the experimenter/operator. It is unclear how the study team accounts for any significant observer/operator bias. Also, for minimally invasive adenocarcinoma (MIA), the tumour can show areas of invasion that appears identical to invasive adenocarcinoma (IAC). It is just that the area of invasion in MIA measures no more than 5mm, while IAC shows invasive areas more than 5mm. Since Raman imaging focused on such small areas (200x200um2), it is unclear to me how this technique can actually distinguish between MIA and IAC without significant guidance from the observer/operator.

Reply 2: Thank you for your valuable suggestions. Firstly, to mitigate operator-induced bias, we have implemented rigorous standardized operational procedures. Prior to selecting the $200 \times 200\ \mu m^2$ analysis areas for Raman imaging, a thorough pathological evaluation is conducted to ensure that the chosen areas are representative of the tissue. Additionally, a double-blind approach is utilized in the selection of collecting areas, whereby the personnel selecting the areas are unaware of the sample's pathological diagnosis, ensuring objectivity and fairness in the analysis. Second, in addressing the challenge of distinguishing between MIA and IAC, we recognized that several small Raman imaging regions on a slice are difficult to accurately distinguish between these two types of adenocarcinomas. Therefore, we consider data from multiple Raman acquisition areas, in conjunction with pathological assessment results. By comparing and analyzing micro-differences in cellular and matrix composition across different areas, we have improved the accuracy of distinguishing between MIA and IAC.

Changes in the text: We have modified our text as advised (see Page 9, Line 225 and Page 19, Line 521), discussing the limitations of this method and suggesting that future research could enhance diagnostic accuracy by increasing the number and/or size of Raman imaging areas and further optimizing algorithms.

Third, 59 cases aren't a lot of cases, but a bigger problem is that very few cases are MIA and how biased the cohort is. In the training set, there were 5 minimally invasive adenocarcinoma (MIA) vs 36 invasive adenocarcinoma (IAC). The ratio of the training set is so biased that it is unclear if the model is reliable. Also, the 6 cases in the authors' test set all represent IAC according to final pathologic diagnosis, so in fact there are no MIA cases in the test set at all. Given the skewed ratio of how IAC is enriched as compared to MIA in both the training set and test set, the seemingly high accuracy reported does not appear reliable. Finally, the authors have used a training set and a test set; however, there is no separate independent validation set to validate the performance of the models. The seemingly high numbers from the ROC's may
likely be due to over-fitting of the models.

Reply 3: We are sorry that we didn't check the model part carefully and there are some mistakes which would lead to misunderstanding. In fact, we adopt balanced weight and 5-

fold cross-validation strategies when training all SVM and CNN models. When training model to distinguish MIA and IAC, we divide 5 MIA and 35 IAC cases into 5 folds (each fold consists of 1 MIA and 7 IAC cases). For each fold, we train a model on the other 4 folds (training set) and evaluate the model on this independent fold (validation set). Therefore, there are actually independent validation sets and the high performance in text is actually the evaluation result which was written as training result incorrectly.

Imbalanced dataset is always a problem when training a model, thus we adopt balanced class weight strategy which can largely increase the training loss when MIA case is classified incorrectly to improve model's ability to recognize the few MIA cases, which is effective according to the 5-fold cross-validation results. Here, to further evaluate the reliability of our models on smaller datasets, we randomly select 25 IAC cases and divide them into 5 sets (5 cases per set). For each set, we train CNN models on it and all 5 MIA cases with balanced weight and 5-fold cross-validation. The final average AUC of all models is 92.9% ± 4.6%, which demonstrate the feasibility of our proposed methodology to distinguish between MIA and IAC.

Finally, regarding the 6 independent cases in our test set, 3 were initially misdiagnosed as MIA during rapid intraoperative analysis. However, all 6 cases were ultimately identified as IAC by the pathological gold standard, leaving no MIA cases in the test set. This discrepancy mainly stems from our limited collection of MIA data, where we managed to gather only 5 cases. We decide use them all to train models with better performance with 5-fold cross-validation strategy, which wouldn't influence their performance on the 6 test cases. In the future, we hope to add more clinical samples, further advancing our findings. Thank you very much for your suggestions, which will guide us in the future.

Changes in the text: We modify the interpretation of the model in the text (see Page 10, Line 267 and Page 16, Line 432).

Fourth, for the refined peak comparison, this appears to be manual correction/removal of spurious peaks by the experimenters. It is unclear if this was conducted blindly to reduce bias in the experimental data.

Reply 4: In this work, all data preprocessing steps were strictly governed by predefined rules, with rigorous parameter constraints for each phase, as detailed in the manuscript. To minimize potential biases introduced during this process, a double-blind approach was employed, wherein the personnel conducting the data preprocessing were unaware of the specific details and pathological diagnoses of the samples. This not only encompassed the data preprocessing procedures but also included the removal of spurious peaks caused by cosmic rays.

Changes in the text: We have modified our text as advised (see Page 9, Line 241). We thank you again for your suggestions and review of this section.

Fifth, Figure 1A is misleading. Intraoperative frozen section pathology does not take 60 minutes in a well-staffed pathology laboratory. It can often be performed in 10-20 minutes.
Reply 5: We appreciate your review and comments. In a pathology laboratory well-staffed, the intraoperative frozen section pathology examination typically can be completed within 10 to 20 minutes. We acknowledge that this might lead to a misinterpretation among readers regarding the time required for intraoperative frozen section pathology examinations.
Changes in the text: We have made appropriate changes to Figure 1 in the revised text. We thank you again for your valuable input.

Reviewer B

Authors presented an innovative approach integrating microscopic Raman techniques to deep learning techniques to improve intraoperative diagnostic quality and distinguish between micro-invasive lung adenocarcinoma by invasive lung adenocaricnoma. The paper is well written.

Despite the article adopted highly innovative methods, here are some major points the authors should solved:

1) In the manuscript the authors presented an innovative way to support pathological intraoperative diagnosis adopting miscroscopic Raman techniques. During the "standard" workflow, one frozen section from fresh (not-fixed) tissue samples (wedge resections) is analyzed be the pathologist to make an appropriate diagnosis. In the article, despite the Figure 1, the "new" proposed workflow results a bit unclear: please, carefully detail every single step of the tissue from the surgical removal to the Raman analysis.
Reply 1: We will modify the Figure 1 in the revised manuscript to include detailed descriptions of each step, encompassing the entire workflow from the surgical excision of tissue samples to the Raman spectroscopy analysis. We appreciate your valuable suggestion, which will help us enhance the clarity and comprehensibility of our article.
Changes in the text: We have made appropriate changes to Figure 1 in the revised text.

2) The main limitation of the standard intraoperative frozen section stands in the inability to evaluate the whole tumor lesion. Some tissue should be preserved for the post-operative evaluation on FFPE samples and molecular analysis. How the new proposed technique can overcome these limitation?
Reply 2: Thank you for your questions in response to our article. One of the core advantages of microscopic Raman spectroscopy is its high specificity for different biomolecules, which

makes it possible to collect detailed information directly from tissue sections with short acquisition times and fast resolved data. Moreover, we performed serial slices of the tissue blocks so that they were in the same coordinates, and when performing Raman data acquisition, we cross-referenced the images against the gold standard of pathology, with the expectation that intraoperative diagnostic accuracy matching the gold standard of pathology would be achieved. This is an advantage over rapid pathology (which is inaccurate), and postoperative pathology gold standard (which does not meet the requirements for intraoperative diagnosis).

3) Please, use a graphical map to represent all patients enrolled and their assignment to the test and training set.

Reply 3: Thank you very much for your suggestion, we will add charts to the supplementary information table S1 to represent all registered patient information.

Changes in the text: We have added charts to Supplemental Information table S1 to show all registered patient information.

Here, a minor point:

1) Deep learning techniques are showing promising application in lung cancer Pathology, not only in the diagnostic field but also regarding prognosis, response to the therapy (Deep-Pathomics) or TILs assessment. Please, report some examples to enrich the discussion section.

Reply: We will add relevant information about the application of deep learning techniques in the field of lung cancer pathology in the discussion section, thank you for your valuable suggestions, which will help us to improve the quality and depth of the article.

Changes in the text: We have modified our text as advised (see Page 18-19, Line 492).

Reviewer C

1. Table 2

a. Please add the unit.

742 **Table 2** Classification performance of two classification algorithms for distinguishing

743 normal tissues, LADC, MIA, and IAC in the training and test cohorts

Arithmetic	Classified tissues	Sensitivity	Specificity	Accuracy	F1_score	AUC
Normal tissue & LADC						
SVM	Training, mean±SD	99.3±0.7	69.5±22.6	94.7±4.4	96.9±2.6	97.6±3.5
	Testing	96.6	63.5	93.2	96.2	96.2

Completed as required.

b. Please indicate how these data are presented in table.

742 **Table 2** Classification performance of two classification algorithms for distinguishing

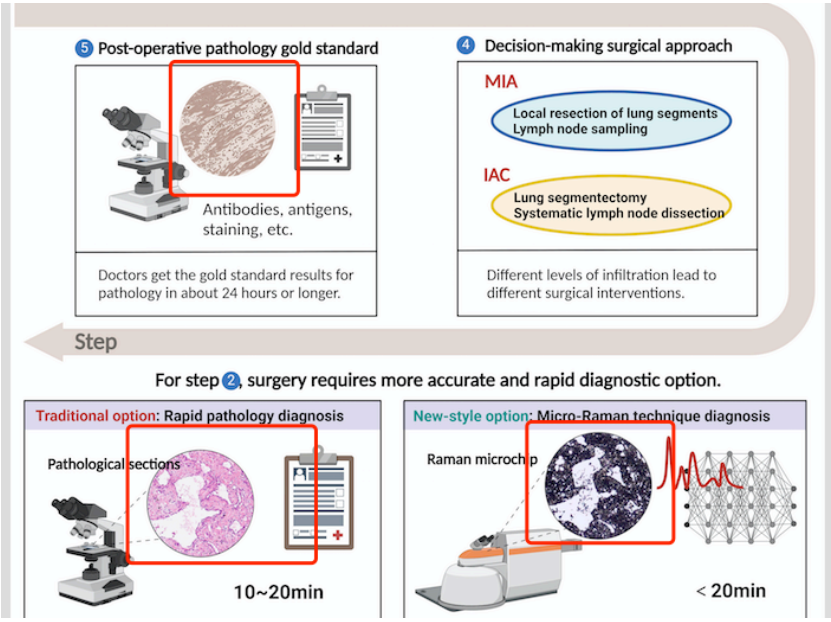
743 normal tissues, LADC, MIA, and IAC in the training and test cohorts

Arithmetic	Classified tissues	Sensitivity	Specificity	Accuracy	F1_score	AUC
Normal tissue & LADC						
SVM	Training, mean±SD	99.3±0.7	69.5±22.6	94.7±4.4	96.9±2.6	97.6±3.5
	Testing	96.6	63.5	93.2	96.2	96.2
CNN	Training	98.0±1.8	92.1±6.4	97.1±1.7	98.3±1.0	99.2±0.9
	Testing	93.3	91.2	93.1	96	96.1
MIA & IAC						
SVM	Training	99.5±0.4	45±28.8	89.2±6.5	93.7±4.0	94.1±9.8
CNN	Training	98.5±1.9	95.8±4.5	98.2±2.0	98.9±1.1	99.4±0.9

We made changes to the table, and in addition we found that the table was more accurately stated using 'validation' instead of 'training'.

2. Figure 1

For cell map, please indicate the magnification (or scale bar) in figure legend.



Completed as required.

3. When using abbreviations in table/figure or table/figure description, please mention the entire expression in a footnote below the corresponding table/figure. Please check and revise. Such as: TP, FP, PN, TN (table S2); LADC, MIA, IAC (table S3).

Completed as required.

4. Reference/citation

The authors mentioned "studies...", while only one reference was cited. Change "Studies" to "A study" or add more citations. Please revise. Please number references consecutively in the order in which they are first mentioned in the text.

Studies have reported 15.6% discrepancy between intraoperative rapid pathology and postoperative gold standard results (13), underscoring the profound impact of these discrepancies on the management of LADC patients, given the distinct clinical significance between minimally invasive and invasive cases.

Completed as required.