

Peer Review File

Article information: <http://dx.doi.org/10.21037/tlcr-20-708>

域代码已更改

Reviewer A:

Comments:

In this interesting review, the authors summarize artificial intelligence in lung cancer screening. It is well written and informative, I found particularly useful the tables of publications. I have only some concerns that can be easily addressed. First, the title mentions screening but the review involves methods used also to characterize the lung tumor later in the patient management. Moreover, the paper focuses too much on radiomics rather than artificial intelligence (machine learning and deep learning).

Comment 1: *A proper definition of Artificial Intelligence should be introduced.*

Reply: Section with a definition of Artificial Intelligence concept has been added

Comment 2: Page 4: *"Application of every mentioned algorithm for lung nodule detection results in a number of false positive nodule candidates. " Please be more precise: how many are the false positives?*

Reply: An additional column with false positive ratio values for each method has been added to Table 2. Unfortunately, some publications do not provide FPR values for their methods.

Comment 3: Page 5 and 6: *the fourth phase of model development is validation (cross validation, external validation...)*

Reply: An additional section describing the fourth phase has been added.

Comment 4: Page 6: *is it possible to give a specific reference to each feature type?*

Reply: References for each feature type have been added

Comment 5: Page 7: *all the section on feature stability would fit well in a radiomics paper but this is an artificial intelligence paper, and it should focus on stability of machine learning models not of features. Consider citing papers on comparison of machine learning models for lung cancer*

Reply: An additional section, showing research on model stability and performance has been added.

Comment 6: *The "Despite this drawback radiomics is still widely used" part is not useful and can be removed*

Reply: This part has been removed.

Comment 7: *"In spite of the importance of radiomics, it is not the only machine learning based techniques that have experienced rapid growth in recent years." is an unclear sentence, rephrase*

Reply: The sentence has been rephrased.

Comment 8: *Consider showing some image of a lung nodule (e.g. one benign and one malignant) and comment it, this would be useful for the reader who is less used to these images.*

Reply: Figure 1 has been added with the depiction of benign and malignant nodule.

Comment 9: *consider citing the following recent review:*

<https://pubmed.ncbi.nlm.nih.gov/32367456/>

Reply: The paper has been added as a reference in the text.

Reviewer B:

Comments:

I read the review paper and I find it generally informative and well-written. Without an attempt to summarize the contents, I will only indicate areas which will profit from improvement.

Comment 1: *The paper includes several tables, in which algorithms and methods of automated detection are compared. There is a difficulty stemming from the fact that each of these uses different quantitative measure of goodness. Table 1 uses “Quality index”, Table 2 uses “Accuracy”, Table 3 uses “Reported sensitivity” and “False positive rate” (is sensitivity the same as true positive rate, if yes, then why can it be not named this way?), Table 4 uses “Sensitivity”, Table 5 uses “Error rate”. In a review paper, comparisons are best standardized and if this is not possible, some explanation is welcome. A good way of doing this is a separate section or a paragraph (or a “box”), explaining reasons for using specific measures of goodness.*

Reply: We unified the nomenclature in reporting measure of goodness. Unfortunately, the differences in performance metrics stem from the fact, that the papers did not publish a single metric that could be used in all comparisons. Thus, several metrics were necessary.

Comment 2: *In at least one place (lines 170–171), the authors discuss gains from early detection of lung cancer. They write “National Lung Screening Trial, proved the effectiveness of radiomics in the early detection of malignant lung nodules, thereby reducing mortality rate by 20% (56);”. What actually happened was that the number of LC-related deaths in the screening group was 20% less than in controls. This is not reduction of mortality, since mortality is a very specifically define epidemiological measure, which does not apply to the situation at hand. Please re-formulate the statement.*

Reply: We apologise for this mistake. This sentence has been rephrased.