

Peer Review File

Article information: <https://dx.doi.org/10.21037/jtd-21-1342>.

Reviewer A:

Comment 1: There is an inconsistency in how the authors report the findings from different studies. For some studies, the statistics of datasets are reported, while for some studies there is limited info on the study cohorts.

Reply 1: We appreciate your comments. As you pointed out, we did not mention all detailed statistics from all studies. There are two reasons for this, one is that some studies did not provide detail information, and the other reason is that we reported detail information only for the studies we consider important. As you advised, this may seem inconsistent, so we have added detailed information about the papers we consider important and put them together in the Table 1.

Comment 2: Also in accordance to #1, it is not clear if the authors had considered any concrete criteria to qualify the studies for this review based on experiments setups. For instance, it is not clear if any numerical limits on the size of subject cohorts had been considered. If there was such criteria, authors should include this info. If not, I would encourage the authors to apply such criteria and clearly mention it.

Reply 2: Thank you for your sharp review. As your comment, numerical criteria should be considered, but it is difficult to apply it to some early-stage AI studies between 2015-2017, since they have preliminary and relatively poor experimental design. Therefore, as an alternative, we decided to separately select good recent studies and organize them in Table 1.

Comment 3: It is unclear for many of the reported studies, if appropriate experiment design with regard to reproducibility and generalizability as well as training, internal validation, and external testing cohorts had been considered. This has been briefly discussed in page 14, lines 306-311 and page 18, lines 379-386. However, this is an extremely critical point in the analysis of AI-based models. Thus, I encourage the authors to revise the review accordingly and ignore studies with poor designs.

Reply 3: Thank you for your comments. We also think that appropriate experiment design such as external validation should be considered. But, it is difficult to exclude all papers without external validation, because most of the important studies in the early stage of medical AI research did not implement external validation. Therefore, rather than excluding all important early studies, we thought it would be better to show good studies in a separate table, so we put them together in Table 1.

Comment 4: To better navigate the reader, it would be great to include a table summarizing the reported studies with regard to applications, cohorts, algorithms, performance, etc. The table should include a row for each study and columns such as publication year, journal, application domain (disease type, etc.), cohort sizes, algorithm used (CNN, U-NET, etc), performance results, reference, ...

Reply 4: Thank you for your sharp review. As your comment, we have added a table summarizing important studies for better navigation (Table 1).

Comment 5: Radiomics features is a hot topic in which AI and DL can contribute extensively. I encourage the authors to add a track for this and report relevant studies focusing on integration of AI and DL methods with radiomics features from different imaging modalities in thorax region. I believe this can further improve the significance of

this review.

Reply 5: We appreciate your comments. As your comment, we have added some studies about radiomics feature and DL applications at the end of the "Automatic quantification for complex quantitative analysis" section of the manuscript (Page 9, line 199-206).

Minor

Comment 6: It would be nice to add a new paragraph at the end of the Introduction part describing what will be discussed in each section in the remaining parts.

Reply 6: Thank you for your suggestion. We added summary of what will be discussed in each section at the end of the "1. Introduction" part (Page 4, line 76-84).

Comment 7: The structure of the document is hard to follow. It would be nice if the sections get ordinal numbers as well.

Reply 7: Thank you for your suggestion. We added ordinal numbers for the main sections.

Comment 8: As stated in page 19, line 413, all the figures are original. However, it is not clearly stated which CAD software (including versions and proper citation of the producer) was used to take the screenshots and which customization criteria are applied. The detailed info should be added accordingly.

Reply 8: Thank you for your suggestion. We added some more detailed information about the CAD software in the captions (Figure 2 (A) and Figure 3 (A)).

Reviewer B:

Comment: This review is very well written. I do not have major recommendations but I suggest to first discuss AI-based triage and worklist prioritization before application of AI

as a second reader

Reply: Thank you for your suggestion. We have re-organized our “2. Application schemes of AI tools in clinical practices” part as advised (Page 5-10).

Reviewer C:

Comment 1: Multiple grammatical errors reduce the manuscript readability, the authors may wish to consider engaging a language editor for this purpose.

Reply 1: Thank you for your suggestion. We have taken a professional language editing service as advised.

Comment 2: While the authors attempted to describe their search criteria in lines 78-83, this is not comprehensive. The number of publications found should be described and reasons for inclusion and exclusions for purposes of the narrative review should be explained. This is to ensure that the article selection is well-balanced, especially since I note that the authors have self-cited a significant number of their previous publications in this review. I suggest that the authors improve on the description of their methodology and revise the manuscript to include any potential publications which have not been included prior.

Reply 2: Thank you for your sharp review. We performed a PubMed literature search and got over 3,600 articles. So, we decided to select best fit articles based on the topic that we choose: AI-based triage and worklist prioritization, AI as a second reader, and Automated quantification for complex quantitative analysis. We first selected papers with a high impact factor and number of citations and looked at the papers that were mainly referenced in these papers. Mentionable topics discovered through this process were used as topics for later sections. It could be confusing, so we have slightly modified it (Page 4, line 73-75).

Also, we have added some potential publications about radiomics which have not been included prior (Page 9, line 199-206).