

Response to Reviewers

Manuscript NO.: 63331, Review

"Radiomics and machine learning applications in rectal cancer: current update and future perspectives"

World Journal of Gastroenterology

We would like to express our gratitude to the Reviewers for their kind remarks and insightful comments that have helped us improve our manuscript. The reviewer queries are addressed in the following point-by-point response and the manuscript has been modified accordingly.

Sincerely,

The Authors

Reviewer #1 Specific Comments to Authors:

- 1) Check the order of the used abbreviations throughout the manuscript. For example, DWI, LR, LARC, ... are used without explaining first.

We apologize for these inconsistencies that have been solved in the revised version of our manuscript.

- 2) Can the authors discuss which sensitivity/specificity should be reached in order to be able to translate these techniques safely into a clinical setting? According to their own opinion or existing guidelines.

The Reviewer raised an interesting point. While high accuracy metrics certainly represent a desirable quality of a predictive model, we believe that they do not represent the main

characteristic to allow a safe implementation in clinical practice. As discussed in the “Current limitations and future perspectives” section, generalizability is probably the key issue since insufficiently tested and robust models might incur severe performance drop when moved into different settings (e.g. different field strengths or vendors for MRI scanner). Furthermore, the value of radiomics-ML models should be demonstrated using current gold standards as benchmarks, their cost-effectiveness should be assessed, and the actual benefit of their implementation explored (e.g. via decision curve analysis). These aspects were also highlighted by Lambin and colleagues in a previous review (10.1038/nrclinonc.2017.141). In the revised version of our manuscript, we have expanded the “Current limitations and future perspectives” section so that it now includes a brief discussion of these issues.

- 3) The expression "the curse of high dimonsionalty" is used. Can this be explained more.

Following the Reviewer suggestion, we have added a brief explanation of the abovementioned expression (which has been slightly modified for simplicity to “curse of dimensionality”) in order to increase the clarity of the manuscript.

- 4) The authors state "higher specificity (95 vs 100%) and specificity (60 vs 43%) compared to those reached by radiologist". Please state more clearly with percentage was obtained by radiologist and which by AI.

We apologize for this typo. Indeed, the numbers reported are correct and thus the AI model had a slightly inferior sensitivity and a higher specificity compared to the radiologist. The sentence has been modified to increase clarity and the typo has been corrected.

- 5) The authors state "although lower values were found in those studies that validated the model in an external dataset". Only 2 studies listed in the table have performed external validation. Therefore no significance can be reached and this statement should be nuanced. For example: a trend can be observed ...

We agree with the Reviewer that the statement was too strong and not sufficiently supported by the collected evidence. Therefore, we modified the sentence following the Reviewer suggestion.

- 6) Within the MRI - genotyping section, prediction of KRAS mutations were discussed. Is there no literature about other mutations/MSS status in RC as has been described in the CT section.

We thank the Reviewer for this insightful comment. To the best of our knowledge, we were only able to find a registered clinical trial that aims to investigate microsatellite instability status in RC using T2w MR images (10.1097/MD.00000000000019428). However, the trial is still ongoing and there are no available results yet. Therefore, we decided not to mention this registered study protocol in our narrative review.

- 7) Can the authors elaborate on which algorithm, as described in table 1, has the highest applicability potential in the clinic (according to their opinion)? Or are they all equivalent?

We thank the Reviewer for this interesting question. All the algorithms presented in table 1 have been object of preliminary investigations in the field of RC radiomics, with promising results. While there are differences among them, there is not one that should be considered intrinsically superior to another. Some could be preferred based on specific characteristics of the employed dataset (e.g. continuous or discrete variables) or of the desired outcome (binary or multiclass classifications). Each has its own advantages (e.g. some are simpler to understand or require less computational power). Therefore, we would not consider them substantially equivalent but at the same time we cannot reliably suggest an superior “all-scenarios fitting” algorithm.

- 8) An additional paragraph describing in short the progress of AI/ML in other cancer types and placing this in the context of RC - lessons to be learned, things in common, most progress in the race to the clinic would be appreciated.

Once again, we appreciate the valid comment made by the Reviewer. Radiomics is a rather young method and its application in the field of oncologic imaging is in the earliest stages. As recently found in a systematic review of the literature, the overall quality of radiomics in oncologic studies is still low regardless the specific cancer setting (10.1007/s00330-019-06360-z). The race is long and RC radiomics is neither ahead nor behind. In the revised version of our manuscript, we have added a sentence supported by the abovementioned study in order to underline this issue.

Reviewer #2 Specific Comments to Authors:

the manuscript is well written and deals with all the current problems in the treatment of rectal cancer. the authors insist in particular on the problem of the therapeutic response to neo-adjuvant treatment and the possibility of organ preservation (looking and waiting). finally, the authors qualify their remarks by evoking the current limits.

We thank the Reviewer for the positive comments.