Peer Review File

Article information: https://dx.doi.org/10.21037/tcr-23-859

Reviewer A

Q1. “No forest plots. This should be the main component in a meta-analysis”
Answer: As the reviewer suggested, forest plots have been added as Figure S1 in the revised manuscript.

Q2. “Improve Figure 2 by adding overall bias”
Answer: As the reviewer suggested, the overall bias has been added in Figure 2 in the revised manuscript.

Q3. “perform meta-analysis comparing studies with and without radiomics inclusion”
Answer: To provide a detailed reference for the selection of variables, we performed a meta-analysis comparing studies with 6 kinds of variables such as radiomics, dosimetric predictor, clinical data, dosimetric predictor + clinical data, radiomics + clinical data, radiomics + dosimetric predictor + clinical data, which was shown in Table 1.

Q4. “No characteristics of the studies”
Answer: The characteristics of the studies, such as authors, last author, publication year, center location, data source, patient inclusion and exclusion criteria, determined criteria of the outcome, time of predictor assessment, and time of outcome assessment, have been provided in Table S3 in the revised manuscript.

Reviewer B

Q1. “In the introduction, it should be stressed treatment strategy, in both endemic and non-endemic countries. Therefore, introduction would be enhanced by addition of references, such as PMID: 27865364 to better contextualize the issue at hand in oncologic scenario.”
Answer: As the reviewer suggested, a more detailed statement of the treatment strategy for Nasopharyngeal carcinoma (NPC) has been added in the “Introduction” part as “Radiotherapy is the mainstay treatment of NPC [1]. For patients with locoregionally advanced NPC, concurrent chemoradiotherapy was also applied [2]”, and the reference PMID: 27865364 has been cited here in the revised manuscript(line 82-83).
[1] Chen YP, Chan ATC, Le QT, et al. Nasopharyngeal carcinoma. Lancet 2019;394:64-80.
[2] Ou D, Blanchard P, El Khoury C, et al. Induction chemotherapy with docetaxel, cisplatin and fluorouracil followed by concurrent chemoradiotherapy or chemoradiotherapy alone in locally advanced non-endemic nasopharyngeal carcinoma.

Oral oncology 2016; 62:114–121.

Q2. “In addition incidence data should be updated, as refers to 2012 (line 52)”
Answer: As the reviewer suggested, incidence data has been added in the “Introduction” part in the revised manuscript as “The age-standardized incidence of NPC was 0.4 per 100 000 to 3.0 per 100 000 [1]”(line 79-80).
[1] Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68:394-424.

Q3. “Text needs some editing (for instance see the first line in key findings, page 3)”
Answer: I’m so sorry we had a text error here. The first line in key findings was modified to “This study presented evidence-based data for the first time on ML models predicting RTLI in NPC patients” in the revised manuscript. As the reviewer suggested, the whole manuscript has been revised to correct grammar, spelling, and syntax errors. The manuscript has also been revised by a native English speaker to ensure the readability of the paper.

Reviewer C

Q1. “What are the biggest strengths and weaknesses of this research model? What is the biggest problem faced? Suggest adding relevant content.”
Answer: As the reviewer suggested, relevant content has been added to the “Introduction” part of the revised manuscript as “The biggest strength of this kind of model is the personalized prediction at the early stage of RTLI whereas the biggest weakness is the dependence on imaging data, which may lead to possible additional medical expenses. ML was used to develop predictive models of NPC prognosis and was proven to have great predictive performance [1]. Nevertheless, the predictive accuracy of ML models of RTLI is still debated since it varies from around 0.31 to 0.97, which may be caused by different variables and ML approaches [2,3]. Also, there is a lack of relevant evidence-based information. Therefore, it is essential to conduct a meta-analysis to evaluate the predictive accuracy of ML for RTLI after radiotherapy”(line 101-109)
[1] Wang Y, He Y, Duan X, et al. Construction of diagnostic and prognostic models based on gene signatures of nasopharyngeal carcinoma by machine learning methods. Transl Cancer Res 2023;12(5):1254-1269.
[2] Lin X, Li Z, Chen S, et al. Divergent white matter changes in patients with nasopharyngeal carcinoma post-radiotherapy with different outcomes: a potential biomarker for prediction of radiation necrosis. Eur Radiol 2022;32:7036-47.
[3] Zhao LM, Kang YF, Gao JM, et al. Functional Connectivity Density for Radiation Encephalopathy Prediction in Nasopharyngeal Carcinoma. Front Oncol 2021;11:687127.

Q2. “How to evaluate clinical benefits based on the content of this study? How to select patients for early clinical intervention? Suggest adding relevant content.”
Answer: As the reviewer suggested, relevant content has been added to the “Discussion” part of the revised manuscript as “LR, SVM, and RF were the most common ML methods used to develop models and all performed well in predicting RTLI, which provided a solution for NPC patients that the radiation dose could be reduced individually at the early stage in the patients with high predicted risk of of RTLI” (line 279-283)

Q3. “There are many databases. Why did the author only select PubMed, Web of Science, Embase, and Cochrane Library databases in this study for searching? Please explain the reason.”
Answer: PubMed, Web of Science, Embase, and Cochrane Library databases are the most common databases and include most medical papers. They were also searched in many high-quality meta-analyses and comprehensive searches could be conducted based solely on these databases [1-3].
[1] Early Breast Cancer Trialists’ Collaborative Group (EBCTCG) & Early Breast Cancer Trialists’ Collaborative Group (EBCTCG). Anthracycline-containing and taxane-containing chemotherapy for early-stage operable breast cancer: a patient-level meta-analysis of 100 000 women from 86 randomised trials. Lancet 2023;401(10384):1277–1292.
[2] Akter S, Islam MR, Rahman MM, et al. Evaluation of Population-Level Tobacco Control Interventions and Health Outcomes: A Systematic Review and Meta-Analysis. JAMA network open 2023;6(7):e2322341.
[3] Xu H, Li T, Shao G, et al. Evaluation of neoadjuvant immunotherapy plus chemotherapy in Chinese surgically resectable gastric cancer: a pilot study by meta-analysis. Frontiers in immunology 2023;14:1193614.

Q4. “This study is based on the analysis and summary of the literatures. It is suggested to add clinical experimental research, which may be more meaningful.”
Answer: The lack of clinical experimental research is one of the limits of this study. However, the result of clinical experimental research is hard to present in the short term because observation of the RTLI usually takes several years. So we reported the preliminary result based on the published literature, in the hope of providing evidence-based information about the predictive value of ML for RTLI after radiotherapy as soon as possible. We have added a related statement in the “Discussion” part of the revised manuscript as “Fourth, it would be better to verify if the SVM is the ML method with the highest predictive performance in more clinical experimental research” (line 336-338).

Q5. “The introduction part of this paper is not comprehensive enough, and the similar papers have not been cited, such as “Construction of diagnostic and prognostic models based on gene signatures of nasopharyngeal carcinoma by machine learning methods, Transl Cancer Res, PMID: 37304552”,“Exosomes derived from umbilical cord

mesenchymal stem cells protect cartilage and regulate the polarization of macrophages in osteoarthritis, PMID: 36267713". It is recommended to quote the articles."
Answer: As the reviewer suggested, a related statement has been added in the "Introduction" part of the revised manuscript as "ML was used to develop predictive models of NPC prognosis and was proven to have great predictive performance [1]" (lline 104-105). The article "Exosomes derived from umbilical cord mesenchymal stem cells protect cartilage and regulate the polarization of macrophages in osteoarthritis, PMID: 36267713" is not related to the research content of this study. Could you please confirm the citation recommendation for this article?
[1] Wang Y, He Y, Duan X, et al. Construction of diagnostic and prognostic models based on gene signatures of nasopharyngeal carcinoma by machine learning methods. Transl Cancer Res 2023;12(5):1254-1269.

Q6. "The number of included studies is too small, and it is recommended to increase the number of multicenter and multi-ethnic studies"
Answer: Unfortunately, because of the geographical distribution of the disease, all included studies were conducted in China and multicenter and multi-ethnic studies weren't been found after a comprehensive search. We will continue to pay attention to this field and conduct new meta-analyses with more multicenter and multi-ethnic studies. A related statement has been added in the "Discussion" part of the revised manuscript (line 330-333).

Reviewer D

Q1. "First, in the title I suggest the authors to replace "predictive value" with "predictive accuracy" since the term value is vague and unclear."
Answer: As the reviewer suggested, we have modified the title to "Predictive accuracy of machine learning for radiation-induced temporal lobe injury in nasopharyngeal carcinoma patients: A systematic review and meta-analysis" in the revised manuscript.

Q2. "Second, the abstract needs some revisions. The authors need to present the controversy regarding the accuracy of ML-based models for the prediction of RTLI and explain why a meta-analysis is needed in the background. In the methods the authors need to describe the literature search, the inclusion criteria and data extraction from included studies. In the results, please report the levels of risk of bias of included studies. The conclusion needs to be tone down since the algorithm is only one of the many factors influencing the predictive accuracy such as the selection of predictors."
Answer: We have modified the "Abstract" part as the reviewer suggested in the revised manuscript(line 37-69).

Q3. "Third, the introduction of the main text is poor. The authors did not review the controversy regarding the accuracy of ML-based models for the prediction of RTLI, analyze the potential reasons for the controversy, and explain why a meta-analysis could

help address this issue"
Answer: As the reviewer suggested, relevant content has been added to the "Introduction" part of the revised manuscript as "Nevertheless, the predictive accuracy of ML models of RTLI is still debated since it varies from around 0.31 to 0.97, which may be caused by with different variables and ML approaches[1,2]. Also, there is a lack of relevant evidence-based information. Therefore, it is essential to conduct a meta-analysis to evaluate the predictive accuracy of ML for RTLI after radiotherapy"(line 106-110).
[1] Lin X, Li Z, Chen S, et al. Divergent white matter changes in patients with nasopharyngeal carcinoma post-radiotherapy with different outcomes: a potential biomarker for prediction of radiation necrosis. Eur Radiol 2022;32:7036-47.
[2] Zhao LM, Kang YF, Gao JM, et al. Functional Connectivity Density for Radiation Encephalopathy Prediction in Nasopharyngeal Carcinoma. Front Oncol 2021;11:687127.

Q4. "Fourth, in the methodology of the main text, please clearly define the studies to be included. I do not think RCTs and case-control studies are eligible since they are not correct studies for the development and validation of predictive models. The studies to be included should be studies of the predictive accuracy. Please describe the details of data extraction."
Answer: As the reviewer mentioned, RCTs are not correct studies for the development and validation of predictive models and we apologize for this mistake. However, as for case-control studies, there were some predictive models based on them and their value was nonnegligible [1-3]. So we have removed RCTs but reserved case-control studies in the inclusion criteria of the revised manuscript.
[1] Palmer JR, Zirpoli G, Bertrand KA, et al. A Validated Risk Prediction Model for Breast Cancer in US Black Women. J Clin Oncol 2021;39(34):3866-3877.
[2] Mbizvo GK, Schnier C, Simpson CR, et al. Case-control study developing Scottish Epilepsy Deaths Study Score to predict epilepsy-related death. Brain 2023;146(6):2418-2430.
[3] Vuong K, Armstrong BK, Drummond M, et al. Development and external validation study of a melanoma risk prediction model incorporating clinically assessed naevi and solar lentigines. Br J Dermatol 2020;182(5):1262-1268.

Reviewer E

Q1. In the Methods section, "Study selection", "Data extraction" and "Quality assessment", what are the statistics for disagreement (i.e., intraclass correlation coefficient [ICC], or Cohen's Kappa)? How were the decisions made (i.e., majority vote or purely by the third reviewer)?
Answer: In the "Study selection", "Data extraction" and "Quality assessment" process, cross-validation was used to check if there was a disagreement between the two reviewers. The decisions were made by a majority vote from these two reviewers and

the third reviewer.

Q2. In the Results section, “Study selection”, Figure 1. The number of studies in the identification and screening phases do not add up. From a total of 1208 records (sum of all sources) and the removal of 1018 before the screening, it is unclear how the count suddenly becomes 699 after the title abstract screening. Therefore, please improve the flowchart. Additionally, please specify the reasons and numbers of exclusions in the identification phase under the description of “Records removed for other reasons (n=287)”.
Answer: I’m so sorry there was an unclear description of the identification phase. The only purpose of this phase was to remove the duplicate records and “Records removed for other reasons (n=287)” means artificial removal of records. The flowchart has been modified in the revised manuscript.

Q3. In the Results section, “Characteristics of included studies”,
a. Please provide a table including the first author, last author, year of publication, center location, and reference of the included studies.
b. Please provide a table demonstrating the characteristics of the included studies. This is rather standard practice in meta-analysis and systematic reviews.
c. Considering search strategies, how did the authors end up with only single-center studies in China being included? Could the authors provide any elaboration?
Answer: As the reviewer suggested, a table containing characteristics such as authors, last author, publication year, center location, and reference of the included studies has been added to the revised manuscript(Table S3). We really hoped to include multicenter and multi-ethnic studies for meta-analysis but after a comprehensive search, only single-center studies in China were included. We have carefully confirmed that the search and selection process was appropriate. The possible reason for this is that Chinese scholars have made more attempts to establish RTLI prediction models because of the high incidence rate of NPC in China.

Q4. In the Results section, “Risk of bias assessment”,
a. Table S2 is not readable. Please improve it. A risk of bias summary figure is suggested.
b. Given that 94.44% of models were found a high risk of bias, it would not be acceptable not to conduct a sensitivity analysis.
c. The domains of risk of bias indicate high risk for participants and outcomes. Therefore, it is not arguable to include a characteristic table of all the included studies.
d. In the domains of risk of bias, the analysis was particularly mentioned. Could the authors please explain why an excessive number of participants would be a problem (or what it exactly means in this context)? Another concern is about the model overfitting and optimism (about 10% of the models). Overfitting and optimisms tend to occur when dealing with big data. In conclusion, the combined use of radiomics and clinical data can derive good models. However, radiomic data are usually big data. Do the authors think the risk of overfitting has influenced results and conclusion?
Answer: a. As the reviewer suggested, the readability of Table S2 has been improved

and the summary of bias has been added in Figure 2 in the revised manuscript.
b. Also, sensitivity analysis has been conducted and showed the stability of the meta-analysis. c. The study characteristics such as data source, patient inclusion and exclusion criteria, determined criteria of the outcome, time of predictor assessment, and time of outcome assessment have been added to Table S2.
d. I'm so sorry there was a wrong expression of risk of bias in the analysis domain. "Excessive number of participants" has been modified to "unreasonable number of participants". As for model overfitting and optimism, the proportion of models with overfitting is acceptable (9.72%) and they had only little influence on the results and conclusion of this analysis.

Q5. "In the Results section, "Meta-analysis", in total, there is one table for C-index and one table for Spe and Sen. However, based on my understanding of the inclusion and exclusion criteria, the included models were not one-to-one comparisons of trained and validated models, but rather a mixed of trained-and-validated, trained-only and validated-only models using different data sources. Therefore, the presented table does not reveal the underlying details. It is expected to have at least two tales for each (C-index, and Spe and Sen): 1) a table presenting results from the same studies; 2) a table presenting results from different studies. Alternatively, and minimally, sensitivity analysis should be conducted to ensure results and findings are valid."
As the reviewer suggested, sensitivity analysis has been conducted. And sensitivity analyses excluding any one model included in this meta-analysis yielded results that were consistent with the primary analysis (Figure S2). In other words, heterogeneity from included studies has little impact on the result. Therefore, we only performed one table that showed the meta-analysis result of the models from different studies, which was grouped by variables or ML methods.

Q6. "In the end, it is commendable that you provided the PRISMA 2020 checklist and explicitly pointed out where the corresponding information can be found in the manuscript. However, please assess the quality of reporting and the overall manuscript using evaluation tools such as AMSTAR2 and GRADE."
As the reviewer suggested, we assessed the quality of this meta-analysis using AMSTAR2 and found it was high. We have added a related statement in the "Discussion" part of the revised manuscript (line 329-330).

Q7. "The symbols before the headings (i.e., #) should be removed"
Answer: The symbols before the headings (i.e., #) were added as the format required by the magazine.

Q8. "In the Methods section, 'Inclusion and exclusion criteria', the criteria should be explained in text form rather than using numbered bullet points."
Answer: As the reviewer suggested, we have modified the numbered bullet points to text form in the "Inclusion and exclusion criteria" part of the revised manuscript.

Q9. In the Methods section, “Exclusion criteria”, “p.5, line 98, what is the basis for 10 RTLI patients? Could the authors please provide justification and references?”
Answer: As the reviewer suggested, the reference to this exclusion criteria was added here in the revised manuscript.
[1] Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction model BMJ 2020;368:m441

Q10. In the Methods section, “Exclusion criteria”, “p.5, line 101, what do the authors mean by mature scale?”
Answer: I’m so sorry we have an unclear expression here. We have modified the “mature scale” to “mature predictive model” in the revised manuscript.

Reviewer F

1. Table S3:
Please check study No. 3. It should be reference 22 in your manuscript. But the title is inconsistent with reference 22. Which one is correct?

3	3	10.1002/hed.27200	Magnetic resonance imaging-based radiomics model for predicting radiation-induced temporal lobe injury in nasopharyngeal carcinoma after intensity-modulated radiotherapy	Dan Bao	Dehong Luo	2022	China

22. Bao D, Zhao Y, Li L, et al. A MRI-based radiomics model predicting radiation-induced temporal lobe injury in nasopharyngeal carcinoma. Eur Radiol 2022;32:6910-21.

Answer: I’m so sorry the title of study No. 3 was wrong. We have corrected it and uploaded the revised Table S3(we renamed it “Table S3-revised-R1”).

2. Table S5:
Could you provide Table S5 in editable format (WORD file), not pdf?
Answer: Only the PDF version of the AMSTAR2 instrument (Table S5) can be downloaded from the official website. We tried to directly convert PDF format to Word format, but it would result in garbled code. Could you please keep the PDF format?

3. Table 1:
Please check the below sentence in your main text. It’s “RF model” in your Table 1, not LR.

300 which exhibited acceptable discrimination. For example, LR models performed well
301 in predicting RTLI, with pooled C-index of 0.815 (95% CI: 0.781–0.850) in the
302 training set, and 0.764 (95% CI: 0.710–0.818) in the validation set. We also

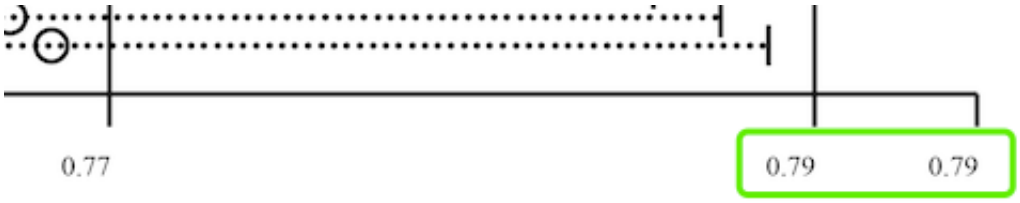
Answer: I'm so sorry for the mistake here. We have modified “LR” to “RF” here in the revised manuscript.

4. Figure S1-2:

1) In your text, Figure S2 was cited earlier than Figure S1. It's not allowed. Please modify the order.

2) Figure S1 is not clear enough. Please provide it in higher resolution.

3) Figure S2: please check whether the data in the x-axis are correct.



Answer:

1) The order of Figure S1 and Figure S2 has been modified in the revised manuscript.

2) We have provided a new figure in higher resolution and renamed it "Figure S2-revised-R1".

3) We found there was a software bug here and uploaded the revised figure(we renamed it "Figure S1-revised-R1").