Manuscript: Imaging Tumor Heterogeneity to Predict Response to Transarterial Chemoembolization in Hepatocellular Carcinoma

Journal: World Journal of Gastroenterology

Manuscript Type: Editorial

I sincerely thank the reviewers for their thorough and constructive critiques. Their comments have strengthened and improved the manuscript.

S.No.	Reviewer's Comments	Response
Reviewer # 1		
1.	This manuscript provides a commentary on the recent work by Lv et al. addressing MRI-based habitat imaging for predicting response to TACE in hepatocellular carcinoma. The topic is of clear clinical interest, particularly given the increasing attention to imaging-derived biomarkers and tumor heterogeneity in treatment decision-making. The author offers a generally thoughtful and well-informed critique, especially with regard to methodological rigor and potential barriers to clinical translation. That said, a few points could be improved to strengthen the manuscript.	Thank you for the reviewer's time and effort to review tge manuscript and for sharing the valuable comments and suggestions to improve the manuscript.
2.	Major Points First, while the manuscript highlights several strengths of the original study, the discussion of its novelty could be made more precise. At present, it is not entirely clear what the author considers to be the key advance—whether it is the use of habitat imaging itself, the ecological diversity metric, or the multicenter validation. A more explicit positioning relative to prior radiomics-based prediction models would help the	Thank you for pointing out. The revised Introduction now clearly identifies the three specific advances introduced by Lv et al. relative to prior radiomics nomograms and AI models

	reader better understand what distinguishes this work.	
3.	Second, a number of methodological limitations are raised (feature selection strategy, treatment heterogeneity, outcome assessment, inter-site variability), all of which are relevant. However, they are presented somewhat uniformly, which makes it difficult to gauge their relative importance. In my view, issues related to potential information leakage during feature selection and variability in outcome definition are likely the most consequential and could be emphasized more clearly, with other points discussed more briefly.	In the revised manuscript, the Feature Selection and Statistical Overfitting section retains its prominent placement, and the language has been sharpened to make clear that information leakage is the single highest-risk methodological concern—it directly threatens the validity of the reported AUC values.
4.	Third, the critique of the statistical approach is generally appropriate, particularly regarding the lack of nested validation and the use of stepwise regression. That said, some of the language is rather definitive. Given that this is an editorial rather than a re-analysis, it might be preferable to frame these concerns more cautiously (e.g., “may introduce bias” rather than “introduces bias”), which would keep the tone balanced.	This is a well-taken point. We have revised the relevant passages throughout the manuscript.
5.	Fourth, the discussion around outcome definition and imaging heterogeneity is useful, especially the point about mixed CT and MRI follow-up. However, parts of this section go into quite detailed methodological alternatives (e.g., modeling outcomes as ordinal or continuous), which may be more than necessary for this format. The core message—potential misclassification and lack of reported inter-reader agreement—is already strong and could stand on its own.	We have reviewed this section, and the differential misclassification risk from mixed CT/MRI follow-up and the absence of inter-reader kappa statistics for mRECIST categorization stand on their own.

6.	Finally, the section on clinical relevance would benefit from being a bit more concrete. The manuscript notes that TACE is increasingly used in combination with systemic therapies, but it remains somewhat unclear how the proposed model might fit into current clinical workflows. A brief discussion of potential use cases (for example, pre-treatment stratification versus early response assessment) and how this approach compares with existing prognostic tools would make this section more practical. Minor Points Some paragraphs, particularly those dealing with statistical considerations, are quite dense and could be shortened slightly for readability.	Thank you for point it out. The revised Clinical Utility and Implementation section now explicitly frames two distinct decision contexts in which this model could plausibly be deployed: (1) pre-treatment stratification, identifying patients unlikely to respond to TACE and redirecting them to systemic therapy upfront; and (2) early response assessment at 4–6 weeks, guiding decisions about repeat TACE versus transition to systemic treatment. We have reviewed the statistical sections for density.
7.	The use of terms such as “generalizability” and “transportability” is somewhat interchangeable; it may be helpful to standardize or clarify their use. • A few statements (e.g., describing certain limitations as “critical”) could be softened to maintain a more neutral tone.	We have reviewed all instances. 'Generalizability' is used in the broader clinical sense of applicability across patient populations and the abstract retains 'transportability' in the technical context where it was originally placed

8.	The reference list is appropriate, although a small degree of consolidation might improve flow. Recommendation Overall, this is a solid and relevant commentary that raises important points about methodology and clinical applicability. With some refinement in emphasis, tone, and clarity, it would be suitable for publication. Recommendation: Minor revision	Thank you for your suggestion. However, references are cited at most relevant point.
Reviewer # 2		
1.	The manuscript titled "Imaging Tumor Heterogeneity to Predict Response to Transarterial Chemoembolization in Hepatocellular Carcinoma" as an editorial has offered a highly critical and necessary evaluation of AI and radiomics applications in oncology. The author has systematically broken down the flaws in retrospective study designs, mixed imaging protocols, and statistical overfitting.	Thank you for the reviewer's time and effort to review tge manuscript and for sharing the valuable comments and suggestions to improve the manuscript.
2.	However, the manuscript needs Major Revisions before publication. There has been a massive block of unformatted bullet points, outline notes, and placeholder tags such as "116364_Auto_Edited-YJP". This section reads like an earlier draft or an AI-generated outline. This entire section should be strictly deleted from the final manuscript, as it breaks the formal editorial structure.	Thank you for pointing out. These pages, an inadvertently appended working document, have been completely removed from the revised submission. The manuscript now concludes cleanly after the reference list.
3.	The phrases have correctly highlighted that the original study lacked a description of missing data extent and imputation strategies.	Thank you for pointing it out and suggestion.

	The authors should add a specific recommendation to the "Study Design and Patient Selection" section suggesting that future studies report missingness patterns and utilize multiple imputation alongside sensitivity analyses.	We have expanded the Study Design and Patient Selection section accordingly. The revised text now explicitly recommends that future studies report missingness patterns for clinical laboratory variables and utilize multiple imputation alongside sensitivity analyses.
4.	The descriptions have emphasized that predicting early response at 4 to 6 weeks lacked justification as a validated surrogate for overall survival or time-to-progression. While the authors have mention this point briefly in the main text, I think the authors should elevate this point to make it clear that without linking the model's predictions to patient-centered endpoints the clinical utility of the habitat model remained purely theoretical.	The revised Clinical Utility and Implementation section now explicitly states that without demonstrating a link between the model's predicted response probabilities and patient-centered endpoints — overall survival, progression-free survival, or time to TACE failure — the clinical utility of the habitat model remains largely theoretical regardless of AUC performance
5.	The authors have critiqued regarding the use of SMDs over P-values for assessing baseline comparability was excellent. It's better to strengthen this by briefly noting that P-values are highly dependent on sample size, making them unreliable for detecting clinically meaningful baseline imbalances in large cohorts.	The revised Study Design and Patient Selection section now explicitly notes that in large cohorts, P-values are highly sensitive to sample size and can achieve statistical significance for clinically trivial imbalances while

		remaining non-significant for meaningful differences in smaller subcohorts. This makes the case for SMDs more mechanistically explicit and directly responsive to the reviewer's reasoning.
6.	In the “Treatment Heterogeneity” section, the authors have appropriately pointed out that treatment decisions, cTACE vs. DEB-TACE were likely based on the same tumor features that affect treatment response. Maybe the authors could mention specific clinical confounders that drive these decisions, for example, Child-Pugh score, ALBI grade, and portal hypertension, which would make this critique even more robust.	Thank you for the suggestion directly strengthens the confounding. The revised Treatment Heterogeneity and Confounding by Indication section now names the specific clinical confounders that drive cTACE versus DEB-TACE selection decisions in practice: Child-Pugh score, albumin-bilirubin (ALBI) grade, portal hypertension, tumor number, and vascular invasion status. We have also expanded the recommendation to call for prospective evaluation of whether model-guided patient selection results in

		meaningfully improved survival outcomes compared to standard clinical decision-making.
Reviewer # 3		
1.	I read this manuscript with great interest. The author tackle a highly relevant clinical challenge: predicting how well HCC patients will respond early on to TACE. To do this, they built a predictive model combining clinical data, conventional radiomics, and MRI-derived habitat imaging (ITH features). The results look impressive at first glance, particularly the high AUC values. That said, I have several major methodological concerns regarding patient selection, feature stability, and how the clinical outcomes were defined. These issues need to be resolved before we can consider this model robust enough for clinical implementation.	Thank you for the reviewer's time and effort to review tge manuscript and for sharing the valuable comments and suggestions to improve the manuscript.
2.	The introduction should clearly compare the proposed approach to existing radiomics nomograms and AI models to highlight what is specifically novel (e.g., habitat construction, ecological diversity index, etc.)	Thank you for the suggestion. The revised Introduction now explicitly compares the Lv et al. approach to prior radiomics nomograms and AI-based models for TACE response prediction, and identifies the advances.
	The authors should acknowledge that MRI-derived	The revised Introduction now

	habitats may reflect a mixture of post-TACE lipiodol effects, perfusion, necrosis, hemorrhage, and acquisition artifacts, rather than purely tumor biology.	explicitly acknowledges that MRI-derived habitat features.
3.	The authors focus exclusively on a "TACE-only" patient population. In current oncology practice, however, TACE is frequently combined with systemic therapies according international guidelines. The manuscript needs to clarify whether this model actually applies to the modern landscape of combination therapies.	The revised Introduction notes that the model was developed in a TACE-only population and that its applicability to the modern treatment, in which TACE is increasingly combined with lenvatinib or immune checkpoint inhibitors, requires explicit consideration
4.	Also, predicting a 4-to-6-week radiographic response is practically useful, but is it a proven surrogate for overall survival or progression-free survival in this specific setting? The introduction should justify this choice. Finally, as radiologists, we should be cautious about attributing MRI habitat features purely to "tumor biology." Post-TACE imaging is notoriously messy.	The Clinical Utility section now explicitly states that the habitat model's clinical utility remains largely theoretical without demonstrated linkage to patient-centered survival endpoints, and calls for prospective evaluation of whether model-guided selection improves outcomes versus standard care.
5.	The features extracted likely capture a mix of lipiodol artifacts, necrosis, perfusion differences, and standard acquisition noise. The authors should acknowledge this limitation early on.	The revised Introduction now explicitly acknowledges that MRI-derived habitat features may reflect not only intrinsic tumor biology, but also a mixture of perfusion differences, post-TACE lipiodol-related effects,

		hemorrhage, necrosis, and acquisition artifacts that this bears directly on the biological interpretability of the features.
6.	Please be aware that the submitted manuscript file contains several pages of what appears to be an internal draft or editorial notes appended after the references (frequently marked with the tag "116364_Auto_Edited-YJP"). You will likely want to clean up and remove these extraneous pages prior to the next round of revisions	These pages, an inadvertently appended working document, have been completely removed from the revised submission. The manuscript now concludes cleanly after the reference list.