120606_Revision_Auto_Edited.docx

Baishideng Publishing Group Co. Limited

Document Details

Submission ID
trn:oid:::3117:580792848

Submission Date
Apr 20, 2026, 4:17 PM GMT+8

Download Date
Apr 20, 2026, 4:20 PM GMT+8

File Name
120606_Revision_Auto_Edited.docx

File Size
21.5 KB

10 Pages

2,259 Words

15,092 Characters

2% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text
- Cited Text
- Small Matches (less than 12 words)
- Abstract
- Methods and Materials

Match Groups

3 Not Cited or Quoted 2%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

2% Internet sources
1% Publications
0% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- **3** Not Cited or Quoted 2%
  Matches with neither in-text citation nor quotation marks
- **0** Missing Quotations 0%
  Matches that are still very similar to source material
- **0** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

2% Internet sources
1% Publications
0% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1 Internet
www.mdpi.com **<1%**

2 Internet
cadth.ca **<1%**

3 Internet
www.qeios.com **<1%**

Name of Journal: *World Journal of Gastroenterology*

Manuscript NO: 120606

Manuscript Type: EDITORIAL

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

Transarterial Chemoembolization in Hepatocellular Carcinoma

Abstract

Magnetic resonance imaging (MRI)-based habitat imaging is a promising non-invasive approach for assessing intratumoral heterogeneity (ITH) and predicting responses to transarterial chemoembolization (TACE) in hepatocellular carcinoma (HCC). A recent multicenter study by Lv *et al*. combined MRI-derived habitat features with clinical data to develop a predictive model that achieved area under the curve values (AUC) of 0.97, 0.91, and 0.93 across different cohorts, outperforming single-modality models. The ecological diversity index from Gaussian mixture model clustering offers a new measure of tumor ecosystem complexity. However, there are methodological concerns, including potential information leakage, confounding factors from mixed TACE methods, misclassification of mRECIST outcomes due to imaging variability, and limited inter-site radiomics harmonization. The low transportability of the standalone clinical model (AUC = 0.54, accuracy = 0.41) underscores significant variability across centers. Future efforts should focus on prospective validation with standardized imaging, integration of multiparametric MRI, multi-lesion modeling, radiogenomic correlation studies, and decision-impact trials to assess whether model-guided patient selection can improve survival outcomes. Addressing these challenges is essential to establishing habitat imaging as a reliable tool for personalized TACE in HCC.

Key Words: Hepatocellular carcinoma; Transarterial chemoembolization; Intratumoral heterogeneity; Habitat imaging; Magnetic resonance imaging; Radiomics; Predictive modeling; Personalized medicine

Core Tip: The recent study by Lv *et al*. integrated contrast-enhanced MRI-derived habitat features, conventional radiomics, and clinical variables to predict early response to transarterial chemoembolization in hepatocellular carcinoma, achieving area under the curve values of 0.97 and 0.93 in the training and validation cohorts, respectively. While the ecological diversity index and multicenter design are meaningful advancements, several key methodological issues remain unresolved: Feature selection

outside nested cross-validation, which risks data leakage and overly optimistic performance estimates; confounding from mixed protocols; outcome misclassification due to the use of both computed tomography and magnetic resonance imaging follow-up; and insufficient inter-site harmonization. Prospective validation, integration of multi-sequence imaging, and studies analyzing the impact on decision-making are necessary before implementing this in clinical practice.

INTRODUCTION

A recent article by Lv JB *et al*.[1], aimed to develop and validate a model that integrates clinical features, radiomics, and intratumoral heterogeneity (ITH) to predict early responses to transarterial chemoembolization (TACE) in patients with hepatocellular carcinoma (HCC). This study makes a significant contribution to personalized oncology by introducing habitat imaging, an innovative technique that quantifies tumor heterogeneity using subregional analysis of magnetic resonance imaging (MRI) data. The combined model demonstrated strong predictive performance, achieving an area under the curve (AUC) of 0.97 on the training set and strong generalizability in both internal (AUC = 0.91) and external (AUC = 0.93) validation cohorts. While prior radiomics nomograms and AI-based models have shown moderate discriminatory ability for TACE response prediction, the key advances introduced by the authors appear to be threefold: The construction of MRI-derived habitat subregions using Gaussian mixture model clustering, the derivation of an ecological diversity index as a novel measure of tumor complexity, and multicenter validation that demonstrates cross-institutional applicability. It is important to note that MRI-derived habitat features may reflect intrinsic tumor biology and a mixture of perfusion differences, post-TACE lipiodol-related effects, hemorrhage, and necrosis, thereby enabling biological interpretation of these features. Additionally, this model was developed in a TACE-only patient population; its applicability to the modern treatment landscape, in which TACE is increasingly combined with systemic therapies such as lenvatinib or immune checkpoint inhibitors, requires further consideration.

Although combining ITH features with clinical data and traditional radiomics is an exciting and relevant approach, several methodological limitations warrant careful attention. These issues pertain to study design, statistical methods, feature selection, and outcome definition, which are crucial for integrating habitat imaging into routine clinical practice. This letter presents the following observations to enhance the interpretation of these findings and guide future research in this promising area.

STUDY DESIGN AND PATIENT SELECTION

The retrospective cohort design introduces selection and survivorship biases that may limit the model's generalizability[2]. The study included patients from two large screening populations: 1,631 from institution A and 743 from institution B. However, it ultimately focused only on those who had (1) a pre-TACE MRI conducted within one month and (2) follow-up imaging 4 to 6 weeks after treatment. This systematic exclusion of patients who experienced rapid deterioration, death, or missing imaging could skew the apparent performance of the model and change the characteristics of the patient population[3].

Additionally, although the authors stated that there were "no significant differences" in baseline characteristics between the cohorts, relying on P-values can be misleading. In large cohorts, P-values are highly sensitive to sample size and can yield statistical significance for clinically minor imbalances, whereas in smaller subcohorts they remain non-significant for meaningful differences, rendering them unreliable for assessing baseline comparability. Using standardized mean differences (SMDs) would provide a stronger indication of baseline comparability, especially for continuous variables such as tumor size, alpha-fetoprotein (AFP), and alkaline phosphatase (ALP) levels[4]. Including a detailed table on missing data and attrition would improve transparency and support a more robust assessment of the study’s external validity. Specifically, future studies in this area should report missingness patterns for clinical laboratory variables and use multiple imputations, alongside sensitivity analyses, to assess whether exclusions due to missing data differ by treatment response.

OUTCOME DEFINITION AND MEASUREMENT VARIABILITY

Using mixed follow-up methods, such as CT or MRI, to assess modified Response Evaluation Criteria in Solid Tumors (mRECIST) can introduce measurement variability, potentially leading to misclassification[5]. The assessment of early arterial enhancement differs significantly between CT and MRI due to differences in temporal resolution, soft-tissue contrast, and sensitivity to small enhancing lesions. This heterogeneity in outcome measurement may systematically affect response categorization, especially in patients with a borderline between partial response and stable disease.

Additionally, inter-reader reliability for outcome assessment was not quantified beyond segmentation inter-class correlation coefficients (ICCs). Since therapy response categories directly influence model training labels, it is important to document inter-reader agreement using kappa statistics or ICCs for mRECIST classification[6]. It is recommended either to limit outcome assessments to a single imaging modality or to stratify analyses by modality when testing for interaction effects. Furthermore, treating outcomes as ordinal categories (complete response, partial response, stable disease, or progressive disease) or modeling continuous tumor changes could provide additional insights.

TREATMENT HETEROGENEITY AND CONFOUNDING BY INDICATION

Regarding treatment heterogeneity and confounding by indication, both conventional TACE (cTACE) and drug-eluting bead TACE (DEB-TACE), with their different bead sizes and chemotherapy protocols, introduce substantial heterogeneity[7]. It is key to recognize that treatments were based on tumor features, liver function, and tumor distribution factors that are also likely to affect treatment response. Specifically, key clinical confounders that drive treatment selection decisions in this setting include Child-Pugh score, albumin-bilirubin (ALBI) grade, portal hypertension, tumor number, and vascular invasion status. Because these same variables influence treatment

response, the model may partially learn treatment-selection patterns rather than capturing tumor biology and its relationship to outcome[8].

To address this limitation, it is recommended to include treatment variables (cTACE vs. DEB-TACE, type of embolic particles, and chemotherapy protocol) as covariates or to perform stratified analyses by TACE type. Using propensity score weighting or matching for TACE type could help disentangle the effects of intratumoral heterogeneity (ITH) features from treatment-related confounding. Also, testing interaction effects (ITH × TACE type) could clarify whether habitat features have different predictive value across treatment modalities.

IMAGING PROTOCOL HARMONIZATION

Multicenter radiomics studies face several challenges due to differences in imaging scanners, acquisition protocols, and reconstruction algorithms[9]. While the discussion covered resampling and normalization procedures, it did not mention specific radiomics harmonization techniques, such as ComBat[10]. Since radiomics features are highly sensitive to acquisition parameters, applying and documenting harmonization procedures would improve confidence in the reproducibility of models across different institutions.

Additionally, exclusive use of arterial-phase images, although convenient, limits the biological information available for assessing heterogeneity. The arterial phase can be influenced by factors such as contrast timing and variations in cardiac output. Incorporating images from other phases, such as the portal venous, delayed, and diffusion-weighted imaging, can improve robustness and biological interpretability[11]. Moreover, performing a site-effect analysis that highlights shifts in feature distribution, model performance across different centers, and calibration drift would provide insights into the model's transportability.

FEATURE SELECTION AND STATISTICAL OVERFITTING

The feature selection process described in the study may introduce significant risks of statistical overfitting and information leakage. The initial dataset includes features, and various filtering techniques—such as variance thresholding, correlation filtering, univariate significance testing, and recursive feature elimination—are applied. However, the lack of rigorous nested cross-validation may lead to overly optimistic assessments of model performance[12,13]. Additionally, conducting extensive univariate testing at a significance level of $P < 0.05$ without adjusting for multiple comparisons may result in false discoveries and unstable findings.

Current radiomics guidelines state that all feature selection should occur within nested cross-validation loops, using only data from the training sets[14]. Using stepwise regression with the Akaike information criterion to select clinical variables may lead to unstable feature selection, biased coefficients, and inflated statistical significance, especially in studies with smaller sample sizes[15]. Instead, penalized logistic regression methods, such as LASSO or elastic net, offer a more reliable approach to variable selection and effectively reduce overfitting.

Although the reported AUC values are notably high, 0.97 during training and 0.91-0.93 in validation, these figures may indicate potential information leakage. To strengthen the validation process, it is advisable to supplement conventional methods with optimism-corrected, bootstrapped AUC estimates and to report calibration slopes and intercepts. The Hosmer-Lemeshow test is discouraged, as it is influenced by sample size and limited in its ability to detect miscalibration[16].

SEGMENTATION STRATEGY

When developing segmentation strategies, it is important to recognize that analyzing only the largest tumor in patients can introduce bias. The total tumor burden and the biological heterogeneity across all lesions significantly influence the overall response[17]. Focusing on the largest tumor might overlook critical insights into the disease dynamics, particularly in cases with satellite nodules or multifocal disease. A thorough approach should either integrate features from all lesions or use metrics such

as maximum, mean, or variance to capture heterogeneity across multiple tumors effectively. Additionally, conducting sensitivity analyses comparing single-lesion and multi-lesion cases can clarify whether the model performs differently across these clinical scenarios. This is particularly relevant since tumor count was identified as an independent predictor in multivariate analyses.

CLINICAL UTILITY AND IMPLEMENTATION

This study focuses on the use of "TACE-only" treatment, which is increasingly being replaced by combination therapies that include systemic treatments[18]. In current clinical practice, TACE is frequently used alongside targeted agents such as lenvatinib and sorafenib, as well as immune checkpoint inhibitors such as atezolizumab and nivolumab, particularly for patients with intermediate- to advanced-stage disease[19,20]. However, the model's relevance to the treatment needs further assessment[21]. It is also important to clarify the intended clinical use of this model: Whether the goal is pre-treatment stratification to identify patients unlikely to respond, or early response assessment at 4–6 weeks to guide decisions about repeat TACE *vs* transition to systemic treatment. Positioning the model within one of these decision contexts and comparing its performance with established TACE prognostic tools, such as the Hepatoma Arterial-embolization Prognostic (HAP) score, within the same dataset would considerably strengthen the clinical applicability of these findings.

Furthermore, while the authors have selected early radiographic response at 4–6 weeks as the primary outcome, this is a clinical approach and lacks validation as a reliable surrogate for long-term outcomes such as overall survival, progression-free survival, or time to TACE failure[22]. This limitation warrants greater importance: Without demonstrating that the model's predictions are associated with patient-centered endpoints, the clinical utility of the habitat model remains largely theoretical, regardless of its AUC. Future studies should prospectively evaluate whether model-guided patient selection improves survival compared with standard clinical decision-making.

Another important yet often overlooked result is the poor performance of the clinical-only model during external validation (AUC = 0.54, accuracy = 0.41), which indicates a substantial decline in performance, representing calibration issues when relying on clinical variables, likely due to differences in patient populations, clinical practices, or outcome measurements across institutions[23]. This finding supports the inclusion of ITH features but warrants further attention.

Although the authors use SHapley Additive exPlanations (SHAP) analysis, providing further clarification would improve its interpretability. Specifically, it would be helpful to know whether SHAP values were derived from the final logistic regression model or from the random forest models, whether they are stable across bootstrap resamples, and how they balance biological relevance with potential objects caused by data acquisition[24].

For practical clinical use, it is important to specify the required inputs, the time and expertise required for segmentation, the availability of software, and adherence to the Image Biomarker Standardization Initiative (IBSI) guidelines[25]. Additionally, comparing the habitat-based model with established TACE prognostic systems, such as the HAP score or the modified HAP score, within the same dataset would provide clinicians with critical context[26].

CONCLUSION

The study by Lv *et al*. presented an innovative application of habitat imaging to predict TACE response in HCC, addressing a significant clinical need for more personalized treatment selection. The integration of ITH features with clinical and traditional radiomics data is theoretically sound and shows promising predictive performance. However, it is essential to maintain methodological rigor in feature selection, validation strategies, and outcome measurement to reliably translate these findings into effective clinical decision support tools[27]. Addressing the methodological considerations outlined will strengthen the evidence base and promote the adoption of habitat imaging in precision oncology for HCC management.

Page 13 of 13 - Integrity Submission Submission ID trn:oid:::3117:580792848