Baishideng® Baishideng Publishing Group

7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA
Telephone: +1-925-399-1568
E-mail: office@baishideng.com
https://www.wjgnet.com

PEER-REVIEW REPORT

Name of journal: *World Journal of Gastrointestinal Oncology*

Manuscript NO: 106103

Title: CT-Based Deep Learning for Preoperative Prediction of Tumor Immune Microenvironment in Colorectal Cancer TSR, TILs, and IS Perspective

Provenance and peer review: Unsolicited Manuscript; Externally peer reviewed

Peer-review model: Single blind

Reviewer's code: 08238819

Position: Peer Reviewer

Academic degree and professional title: PhD

Reviewer's Country/Territory: United States

Author's Country/Territory: China

Manuscript submission date: 2025-02-15

Reviewer chosen by: AI Editor

Reviewer accepted review: 2025-02-17 08:58

Reviewer performed review: 2025-02-23 04:29

Review time: 5 Days and 19 Hours

Scientific quality	Grade B (Very good)
Novelty of this manuscript	Grade B (Very Good)
Creativity or innovation of this manuscript	Grade B (Very Good)
Scientific significance of the conclusion in this manuscript	Grade B (Very Good)
Language quality	Grade C (Good)
Does this manuscript describe a study of	Yes

7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA
Telephone: +1-925-399-1568
E-mail: office@baishideng.com
https://www.wjgnet.com

the existing knowledge system?	
Does this manuscript report a revolutionary innovation?	No
Does this manuscript report an unconventional innovation?	Yes
Conclusion	Major revision
Re-review	Yes
Peer-reviewer statements	Peer-Review: Anonymous
	Conflicts-of-Interest: No

SPECIFIC COMMENTS TO AUTHORS

The manuscript presents a retrospective study employing CT-based deep learning (DL) radiomics for the non-invasive preoperative evaluation of the tumor immune microenvironment (TIME) in colorectal cancer (CRC). The proposed DL models demonstrate promising predictive performance in assessing key immune-related biomarkers, such as the tumor-stroma ratio (TSR), tumor-infiltrating lymphocytes (TILs), and immune score (IS). While the study is scientifically significant, certain aspects require further refinement to improve clarity, reproducibility, and clinical applicability. For Methods: 1. Feature Importance & Model Interpretation The manuscript lacks a detailed analysis of the most influential features contributing to model performance. It would be valuable to include feature importance analysis using SHAP (Shapley Additive Explanations) or other interpretable AI techniques to better understand which radiomics or DL-derived features drive model predictions. 2. Justification for DenseNet-169 Selection While the manuscript compares multiple deep learning architectures, it does not provide a clear justification for why DenseNet-169 was ultimately selected. It appears that AUC values were used for comparison, but the reasoning should be made more explicit. For instance, were other performance metrics

7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA

Telephone: +1-925-399-1568

E-mail: office@baishideng.com

https://www.wjgnet.com

(e.g., sensitivity, specificity, calibration errors) also considered? 3. Confidence Interval Overlaps Although DenseNet-169 achieved the highest AUC, its confidence interval (CI) overlaps with those of other models. This raises questions about whether the difference is statistically significant. The authors should explicitly discuss this overlap and clarify whether the performance superiority is robust or potentially due to sample variation. 4. Segmentation Consistency & Interobserver Variability The CT segmentation and labeling process is well described, but interobserver variability among radiologists is not sufficiently addressed. Reporting inter-rater agreement metrics such as Dice similarity coefficient or Cohen's kappa would help demonstrate annotation consistency and validate the reliability of tumor region delineation. 5. Training vs. Validation Performance Discrepancies In Table 2, some models exhibit better performance in the validation set than in the training set, which is unusual because training typically involves replicating known data and should, in theory, achieve a "perfect" AUC. This discrepancy raises concerns about potential issues in hyperparameter tuning, overfitting, or data leakage. Possible reasons for this observation should be discussed. Moreover, if hyperparameter optimization was conducted, the methodology should be clearly described. For Tables & Figures: 1. Table 1 - Statistical Methods Clarification The table legend should explicitly describe the statistical methods used for categorical and numeric variables to enhance clarity. Additionally, the reference to "a" is missing and should be checked. 2. Formatting in Tables & Text There should be a space between numbers and parentheses throughout the manuscript (e.g., change "2(1.5,3.2)" to "2 (1.5, 3.2)"). This issue is present in multiple sections and should be systematically corrected. 3. Figure 1 - Clarity & Labeling - The resolution (DPI) of Figure 1 should be improved for better readability. - For the deep learning procedure visualization, the X-axis and Y-axis must be clearly labeled to indicate what is being measured. - The authors should explicitly state that subsequent plots (e.g., Grad-CAM visualizations, ROC curves)

7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA
Telephone: +1-925-399-1568
E-mail: office@baishideng.com
https://www.wjgnet.com

correspond to the best-performing model. However, it would be preferable to first showcase model development and comparisons, followed by the performance of the selected model. For Writing & Formatting & Grammar & Typographical Errors: 1. There are several minor typos throughout the manuscript. For example, "CRC is one the main causes…" → should be "CRC is one of the main causes…". 2. In citation formatting, ensure proper spacing after periods (e.g., "[1].It is anticipated…" should be "[1]. It is anticipated…").