Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Pretreatment CT-based clinical deep learning radiomics model for predicting the response of hepatocellular carcinoma patients with portal vein thrombosis treated with TKIs plus PD-1 inhibitors and hepatic arterial infusion chemotherapy

Guorui Wang
Shandong Cancer Hospital and Institute, Shandong First Medical University, Shandong Academy of Medical Sciences
Fang Liu
Shandong Cancer Hospital and Institute, Shandong First Medical University, Shandong Academy of Medical Sciences
Chunxue Wu
Shandong Cancer Hospital and Institute, Shandong First Medical University, Shandong Academy of Medical Sciences
Bin Zhang
Shandong Cancer Hospital and Institute, Shandong First Medical University, Shandong Academy of Medical Sciences
Hao Zhang
Shandong Cancer Hospital and Institute, Shandong First Medical University, Shandong Academy of Medical Sciences
Hongtian Cheng
Shandong Cancer Hospital and Institute, Shandong First Medical University, Shandong Academy of Medical Sciences
Yimin Sun
Shandong Cancer Hospital and Institute, Shandong First Medical University, Shandong Academy of Medical Sciences
Xiaochen Ren
Dongying District People's Hospital
Jiadong Liu
Shandong Cancer Hospital and Institute, Shandong First Medical University, Shandong Academy of Medical Sciences
Jianjun Han
hanjianjun@sdfmu.edu.cn

Shandong Cancer Hospital and Institute, Shandong First Medical University, Shandong Academy of Medical Sciences

Article

Keywords: hepatocellular carcinoma, portal vein tumor thrombus, radiomics

Posted Date: November 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11554224/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

This study aimed to construct a pretreatment contrast-enhanced CT (CECT)-based multimodal model for noninvasive forecasting of early treatment response in hepatocellular carcinoma (HCC) patients with portal vein tumor thrombosis (PVTT) who received hepatic arterial infusion chemotherapy combined with tyrosine kinase inhibitors and PD-1 inhibitors (HTI).

Methods

We retrospectively enrolled 116 patients with HCC and PVTT receiving HTIs (2019–2024), randomly divided 7:3 into training (n = 81) and test (n = 35) cohorts. Early objective response was assessed 6–8 weeks after treatment using modified RECIST. Radiomic features were extracted from three-dimensional volumes of interest of the primary tumor and PVTT; deep features from the largest axial tumor slice using ResNet101. Six models were developed; the final radiomic–deep learning–clinical (RCLD) model integrated out-of-fold component probabilities. Discrimination, calibration, clinical utility, and overall survival (OS) stratification were evaluated.

Results

Sixty-eight patients responded (complete response, n = 3; partial response, n = 65); the objective response rate was 58.6% and disease control rate 91.4%. Four radiomic and three deep learning features were selected; cirrhosis was the only clinical variable. RCLD achieved the highest AUC (training 0.965, 95% CI 0.929–1.000; test 0.924, 95% CI 0.822–1.000). In the test cohort, RCLD outperformed clinical-only and radiomics–clinical models (p = 0.003 and p = 0.026), with no significant differences from other imaging-based models. RCLD stratified OS (training P = 0.003; test P = 0.027).

Conclusion

A pretreatment CECT-based multimodal model integrating radiomics, deep learning, and cirrhosis predicted early objective response and stratified OS risk in HCC with PVTT receiving HTIs. Prospective multicenter validation is required.

Introduction

Primary liver cancer is among the most common malignancies[1], and hepatocellular carcinoma (HCC), the predominant histologic type[2], frequently invades the portal venous system, causing portal vein tumor thrombosis (PVTT), a major poor-prognosis complication[3]; PVTT is one of the most common complications and a significant predictor of poor prognosis[4]. The Western guidelines classify HCC with PVTT as BCLC stage C and recommend systemic therapy, typically sorafenib, but median overall survival (OS) with sorafenib monotherapy remains only 6.5–10.7 months[5–7]; HAIC combined with molecularly targeted therapy and immunotherapy (HTIs) has shown encouraging survival and response benefits, including in PVTT[8]. In a real-world study, HTIs plus lenvatinib and a PD-1 inhibitor prolonged median OS (19 vs. 14 months) and PFS (10 vs. 4 months), and a multicenter propensity-matched study in Vp4 PVTT reported median OS of 21.1 vs. 11.5 months, median PFS of 8.4 vs. 5.6 months, ORR of 60.6% vs. 28.9%, and DCR of 95.4% vs. 77.6%[9]. However marked HCC heterogeneity means that not all patients derive meaningful benefits.

The unresolved clinical problem is how to identify, before treatment initiation, patients most likely to benefit from HTIs[10,11]. Conventional clinical and pathological features, such as tumor size and number, alpha-fetoprotein, and Child-Pugh class, incompletely capture intratumoral heterogeneity and complex biology, whereas radiomics and deep learning (DL) may provide complementary noninvasive information[12–14]. A prior CT radiomics (Rad) model predicted outcomes after lenvatinib, PD-1 inhibitor, and interventional therapy with AUCs of 0.900 and 0.893, but related models have mainly examined broader unresectable HCC populations[14], single local therapies, or regimens differing from HAIC-based HTIs[15–17]; no study has specifically integrated three-dimensional handcrafted CT Rad, DL representations, and clinical factors to predict early HTIs response in HCC-PVTT, a clinically relevant gap because early identification could spare nonresponders ineffective treatment, toxicity, and cost.

We therefore aim to use pretreatment CECT images and baseline clinical characteristics to construct and validate Rad, DL, clinical (Clinic), and multimodal fusion models for early response estimation in patients with HCC-PVTT receiving HTIs. We will further explore their potential value for pretreatment risk stratification, addressing an urgent unmet need in current clinical practice. By combining these complementary data

sources, the models may improve early response prediction and support pretreatment risk stratification in this high-risk population after independent external validation and clinical utility assessment in future studies.

Method

Data preparation and patient selection

This investigation adhered to the principles set forth in the Declaration of Helsinki (2013 revision). The study protocol received ethical approval from the Ethics Committee of Shandong Cancer Hospital and Institute, Shandong First Medical University and Shandong Academy of Medical Sciences (Approval No. 202607051). The requirement for informed consent was waived. Clinical data and pretreatment CECT images were retrospectively reviewed for 116 eligible patients with HCC and PVTT who received HTI at Shandong Cancer Hospital and collaborating institutions from January 2019 through December 2024. The patient inclusion criteria, enrollment process and data preparation are presented in Appendix E1 and Supplementary Figure S1.

Implementation of the treatment and response assessment

For the implementation of the treatment and response assessment, see Appendix E2. Participants who achieved CR or PR were classified into the treatment responder (TR) group, whereas those with SD or PD were classified into the no treatment responder (nTR) group.

(a)CT image acquisition and segmentation. (b) Radiomics feature extraction. (c) Radiomics features

selection. (d) Deep learning feature extraction and visualization. (e) Clinical feature selection and multimodal integration. (f) Classifier comparison and final model construction.

Grad-CAM, gradient-weighted class activation mapping; RF, random forest; ROI, region of interest.

CECT image acquisition and preprocessing

CECT is not only the most widely used diagnostic imaging technique, but also an important method for diagnosing HCC-PVTT and, even more, an important method for evaluating treatment response in patients. CECT image acquisition and preprocessing are presented in the Appendix E3 and Fig. 1a.

Region of interest segmentation, and feature extraction

Figure 1a, b illustrate the process of region of interest (ROI) segmentation and feature extraction in this study. To guarantee accuracy and stability, two experienced radiologists independently delineated all ROIs, with any disagreements finally resolved by a senior professor of radiology. Finally, 1,563 features were extracted via PyRadiomics version 3.0.1. Details of ROI segmentation and feature extraction are provided in Appendix E4.

Rad, Clinic and DL feature selection and model construction

In this study, the 2D ROI with the largest area was used as the basis for deep learning. Using ResNet101 as the backbone, we established a corresponding deep learning model and extracted 2,024 deep learning features, which were then subjected to principal component analysis (PCA) to obtain the seven most important features (Fig. 1e). For Rad and DL construction, the features were analyzed by logistic regression with least absolute shrinkage and selection operator (LASSO) penalization, so as to identify the most representative features (Fig. 1c). For screening of clinical variables, continuous data with a normal distribution were summarized as mean ± standard deviation (SD) and compared between groups by the independent-samples t-test. Continuous data with a skewed distribution were summarized as the median with interquartile range (IQR) and analyzed using the Mann–Whitney U test. Categorical data were summarized as frequencies and percentages and compared by the chi-square test or Fisher's exact test, as appropriate. Categorical variables were expressed as frequencies and percentages and compared using the chi-square test or Fisher's exact test, as appropriate. Candidate clinical variables were evaluated using univariable logistic regression (Fig. 1e). The detailed contents of the Rad and DL feature selection section are presented in Appendix E5.

Gradient-weighted class activation mapping (Grad-CAM) was used to visualize the DL model by weighting the feature maps of the Convolutional Neural Network's final convolutional layer according to their gradient information[18]. The resulting heatmaps highlighted the image regions that contributed most strongly to the model's predictions, thereby enhancing the visualization and interpretability of the model's decision-making process.

The Rad and DL models were each developed using five machine-learning algorithms: random forest (RF), support vector machine (SVM), Light Gradient Boosting Machine (LightGBM), Adaptive Boosting (AdaBoost), and k-nearest neighbors (KNN). The clinical model was constructed using logistic regression (LR) (Fig. 1f).

Fusion model construction and classifier selection

Three multimodal fusion models were built using probability-level late fusion: DL-Rad (DLR), Rad-Clinic (RDC), and Rad-Clinic-DL (RCLD). SVM, RF, and LightGBM were compared by stratified K-fold cross-validation using AUC, accuracy, sensitivity, specificity, and F1-score. Details on fusion model construction and classifier selection are provided in the Appendix E6.

Response Risk Nomogram and survival analysis

In this study, we established a response nomogram that incorporated cirrhosis, the Rad score, and the DL score; the RCLD score stratified OS using a training-derived cutoff, with Kaplan–Meier and log-rank analyses. Details of the response risk nomogram and survival analysis are provided in Appendix E7.

Statistical analysis

Python (version 3.9) was used for constructing the prediction models. Statistical analyses were carried out with R software (version 4.3.0, http://www.R-project.org) and SPSS (version 25.0, IBM). Continuous variables were analyzed by the independent-samples t-test or the Mann–Whitney U test, whereas categorical variables were examined using the Pearson χ^2 test or Fisher's exact test. AUC values were compared using the Delong test[19]; clinical utility was assessed via DCA and validated in the validation set. For survival analysis, patient OS was assessed using Kaplan–Meier curves, and differences between the two groups were examined with the log-rank test. In addition, ROC curves and calibration curves were generated in the R environment to assess the performance of the prediction models. Statistical significance was set at a two-sided $p < 0.05$.

Results

Clinical characteristics and treatment response

A total of 212 patients with HCC and PVTT who received HTIs were retrospectively screened. After excluding 96 patients according to the prespecified eligibility criteria, 116 patients were ultimately enrolled. Among them, 81 patients constituted the training cohort and 35 constituted the test cohort. The detailed patient selection process and reasons for exclusion are presented in the study flowchart.

Among the 116 enrolled patients, 68 (58.6%) had TR, whereas 48 (41.4%) had nTR. Accordingly, the overall ORR was 58.6%. The detailed distribution of the tumor response was as follows: CR, n = 3; PR, n = 65; SD, n = 38; and PD, n = 10, resulting in a DCR of 91.4%.

The baseline characteristics of the patients in the training and test cohorts are summarized in Table 1. Most demographic, etiological, liver function, laboratory, tumor-related, and staging variables were similarly distributed between the two cohorts. However, a statistically significant difference in AST level was observed between the training and test cohorts ($p = 0.017$). No statistically significant differences were found in the remaining baseline variables (all $p > 0.05$).

Table 1
Baseline characteristics of patients in training and test cohorts.

Variable	Overall (N = 116)	Training (N = 81)	Test (N = 35)	P-value
Demographic				
Age, years, mean ± SD	56.23 ± 10.63	55.68 ± 10.88	57.51 ± 10.05	0.491
Height, cm, mean ± SD	170.58 ± 7.10	170.91 ± 6.74	169.80 ± 7.91	0.440
Weight, kg, mean ± SD	70.37 ± 11.91	70.74 ± 11.84	69.52 ± 12.21	0.547
BMI, kg/m², mean ± SD	24.11 ± 3.38	24.14 ± 3.24	24.05 ± 3.74	0.891
Sex, n (%)				1.000
Female	14 (12.1)	10 (12.3)	4 (11.4)	
Male	102 (87.9)	71 (87.7)	31 (88.6)	
ECOG, n (%)				0.220
0	82 (70.7)	54 (66.7)	28 (80.0)	
1	34 (29.3)	27 (33.3)	7 (20.0)	
Etiology & Liver Function				
Etiology, n (%)				
Hepatitis B virus	102(87.93)	72(88.89)	30(85.71)	
Hepatitis C virus/Other	14(12.07)	9(11.11)	5(14.29)	
HBV DNA, n (%)				0.842
Negative	63 (54.3)	43 (53.1)	20 (57.1)	
Positive	53 (45.7)	38 (46.9)	15 (42.9)	
Liver cirrhosis, n (%)				0.271
No	33 (28.4)	26 (32.1)	7 (20.0)	
Yes	83 (71.6)	55 (67.9)	28 (80.0)	
Child-Pugh, n (%)				0.397
A	85 (73.3)	57 (70.4)	28 (80.0)	
B	31 (26.7)	24 (29.6)	7 (20.0)	
ALBI score, mean ± SD	-2.62 ± 0.80	-2.60 ± 0.85	-2.68 ± 0.70	0.420
Laboratory				
ALT, U/L, mean ± SD	53.98 ± 51.60	59.12 ± 58.36	42.10 ± 28.01	0.121
AST, U/L, mean ± SD	85.67 ± 95.58	95.87 ± 109.83	62.07 ± 41.21	**0.017**
AST/ALT ratio, mean ± SD	1.83 ± 1.19	1.92 ± 1.27	1.63 ± 0.97	0.280
Albumin, g/L, mean ± SD	40.84 ± 8.97	40.54 ± 9.33	41.53 ± 8.17	0.390
TBIL, μmol/L, mean ± SD	22.34 ± 14.68	22.65 ± 16.25	21.62 ± 10.30	0.698
PT, s, mean ± SD	12.46 ± 1.34	12.48 ± 1.34	12.41 ± 1.35	0.990
AFP, ng/mL, mean ± SD	37470.29 ± 110261.20	32148.72 ± 94779.11	49785.92 ± 140573.14	0.754
Tumor Characteristics				
Tumor diameter, cm, mean ± SD	9.49 ± 3.98	9.73 ± 4.01	8.94 ± 3.92	0.325
Tumor number, n (%)				0.686
Single	48 (41.4)	35 (43.2)	13 (37.1)	

Variable	Overall (N = 116)	Training (N = 81)	Test (N = 35)	P-value
Demographic				
Multiple (≥ 2)	68 (58.6)	46 (56.8)	22 (62.9)	
PVTT classification, n (%)				0.939
1–2	39 (33.6)	27 (33.3)	12 (34.3)	
3	29 (25.0)	21 (25.9)	8 (22.9)	
4	48 (41.4)	33 (40.7)	15 (42.9)	
Staging				
T stage, n (%)				0.885
2	5 (4.3)	3 (3.7)	2 (5.7)	
3	51 (44.0)	36 (44.4)	15 (42.9)	
4	60 (51.7)	42 (51.9)	18 (51.4)	
N stage, n (%)				0.058
0	96 (82.8)	63 (77.8)	33 (94.3)	
1	20 (17.2)	18 (22.2)	2 (5.7)	
M stage, n (%)				0.742
0	109 (94.0)	77 (95.1)	32 (91.4)	
1	7 (6.0)	4 (4.9)	3 (8.6)	
TNM stage, n (%)				0.284
2	5 (4.3)	4 (4.9)	1 (2.9)	
3	60 (51.7)	38 (46.9)	22 (62.9)	
4	51 (44.0)	39 (48.1)	12 (34.3)	

BMI, body mass index; PVTT, portal vein tumor thrombus; PT, prothrombin time; TBIL, total bilirubin; ALT, alanine aminotransferase; AST, aspartate aminotransferase; AFP, alpha fetoprotein; ALBI, albumin-bilirubin; ECOG, Eastern Cooperative Oncology Group.

Clinic, Rad, and DL model construction and validation

Candidate clinical variables were first examined using univariable logistic regression. Variables meeting the prespecified screening criterion were considered for multivariable analysis. Liver cirrhosis was the only eligible variable (OR = 1.750, 95% CI 1.104–2.776; $p = 0.046$), while all remaining variables had $p > 0.05$ (Table 2). The planned multivariable logistic regression therefore reduced to a single-predictor Clinic model containing cirrhosis. Because no other covariates entered the model, the reported estimate was identical to the univariable estimate and should not be interpreted as a fully adjusted independent association.

Table 2
Univariate and multivariate logistic regression analyses of clinical characteristics for TR in the training cohort.

feature_name	OR_UNI	OR lower 95%CI_UNI	OR upper 95%CI_UNI	*p*_value_UNI	OR_MULTI	OR lower 95%CI_MULTI	OR upper 95%CI_MULTI	*p*_value_MULTI
M_stage	0.333	0.05	2.228	0.341				
ALBI	0.905	0.791	1.037	0.228				
ECOG	0.929	0.493	1.749	0.847				
AST	0.999	0.997	1.002	0.617				
N_stage	1	0.461	2.171	1				
AFP	1	1	1	0.281				
High, cm	1.001	0.999	1.003	0.327				
ALT, U_L	1.001	0.996	1.005	0.847				
Age, year	1.003	0.997	1.01	0.396				
Wight, kg	1.004	0.998	1.009	0.256				
TBIL, μmol_L	1.004	0.991	1.018	0.583				
Albumin, g_L	1.006	0.997	1.015	0.254				
BMI, kg_m2	1.011	0.996	1.026	0.245				
PT, s	1.017	0.987	1.047	0.356				
Tumor_diameter, cm	1.023	0.988	1.06	0.279				
AST/ALT	1.046	0.891	1.228	0.643				
TNM stage	1.055	0.949	1.172	0.407				
T stage	1.074	0.968	1.192	0.258				
BCLC stage	1.077	0.953	1.218	0.318				
PVTT classification group	1.079	0.961	1.21	0.282				
HBV DNA	1.111	0.651	1.896	0.746				
reason	1.118	0.758	1.649	0.638				
Tumor number	1.148	0.918	1.436	0.312				
Chlid-pugh	1.187	0.907	1.553	0.295				
Sex	1.221	0.89	1.674	0.298				
cirrhosis	1.75	1.104	2.776	0.046	1.75	1.104	2.776	0.046

OR, odds ratio; CI, confidence interval; The bold values mean the *p* value < 0.05

A total of 1,563 handcrafted radiomic features were extracted from each three-dimensional ROI. Low-variance filtering, Z-score normalization, independent-samples *t* test, Spearman correlation filtering, and LASSO logistic regression with 10-fold cross-validation retained four features: gradient_firstorder_Minimum, wavelet_HLL_glrlm_GrayLevelVariance, wavelet_HHL_ngtdm_Busyness, and wavelet_HLL_ngtdm_Strength. SHAP analysis ranked gradient_firstorder_Minimum as the largest contributor to the Rad, followed by GrayLevelVariance, Busyness, and Strength. Their SHAP values extended to both sides of zero, indicating that these features contributed to individual predictions with different directions and magnitudes.

For DL analysis, the global average pooling layer of the pretrained ResNet101 generated 2,024 features per CT image. Principal component analysis reduced these features to seven components, and the three components most closely associated with treatment response were retained. RF, SVM, KNN, LightGBM, and AdaBoost were compared by cross-validation for both the Rad and DL models. RF provided the best overall cross-validation performance and was selected to construct Rad-RF and DL-RF. All preprocessing, dimensionality reduction, and feature-

selection procedures were fitted in the training cohort and applied unchanged to the internal test cohort. Grad-CAM visualizations were also generated to identify and compare the model-attention patterns between the TR and nTR groups.

All preprocessing, dimensionality reduction, and feature-selection procedures were fitted in the training cohort and applied unchanged to the test cohort. For the Clinic, Rad-RF, and DL-RF models, the training-cohort AUCs reached 0.611 (95% CI 0.508–0.714), 0.906 (95% CI 0.844–0.969), and 0.815 (95% CI 0.719–0.911), respectively(Table 3). Thus, the Rad and DL models showed similar discrimination in the test cohort and both had higher AUC point estimates than the clinical model.

(a, b) ROC curves in the training (a) and test (b) cohorts, with the RCLD model achieving the highest AUC. (c, d) Decision curve analyses of the training (c) and test (d) cohorts. (e, f) Calibration curves showing agreement between predicted and observed probabilities. (g, h) Kaplan–Meier curves stratified by the RCLD score, showing poorer OS in the high-risk group.

AUC, area under the curve; DCA, decision curve analysis; OS, overall survival; ROC, receiver operating characteristic; SVM, support vector machine SVM, support vector machine; RCLD, radiomics-deep learning-clinical model.

RAD, DLR and RCLD model construction and validation

We constructed three multimodal fusion models by late fusion at the decision level. Specifically, the RDC model merged the probability outputs of the Rad and Clinic component models; the DLR model merged those of the Rad and DL component models; and the RCLD model merged the outputs of the Rad, DL, and Clinic component models. The DLR model was generated by late fusion of the predicted probabilities from the Rad and DL component models. The RCLD model was generated by late fusion of the predicted probabilities from the Rad, DL, and Clinic component models. SVM, RF, and LightGBM classifiers were compared using stratified cross-validation within the training cohort. RF achieved the best cross-validation performance for RDC and DLR, yielding RDC-RF and DLR-RF, whereas SVM was selected as the meta-learner for RCLD-SVM. Out-of-fold component probabilities were used to generate the late-fusion models and to train the meta-learner, thereby reducing information leakage. In the training cohort, the AUCs of RDC-RF, DLR-RF, and RCLD-SVM were 0.948 (95% CI 0.907–0.989), 0.931 (95% CI 0.883–0.980), and 0.965 (95% CI 0.929–1.000), respectively. Their corresponding AUCs in the test cohort were 0.850 (95% CI 0.714–0.985), 0.855 (95% CI 0.717–0.993), and 0.924 (95% CI 0.822–1.000) (Fig. 2a, b). RCLD model therefore had the highest AUC point estimate in both cohorts. These findings suggest that the hybrid model outperformed all other models in predictive performance. Combining radiomic and deep learning-derived features with clinical information further enhanced its predictive accuracy. DCA demonstrated that RCLD conferred a larger net benefit than the alternative models over the relevant range of threshold probabilities in every cohort examined (Fig. 2c, d). In addition, Calibration curves for all models showed that the RCLD predicted TR probabilities were in good agreement with the observed values (Figure. 2e, f). DeLong analysis showed that although the RCLD model achieved statistical significance ($p < 0.05$) only versus the Clinic and RDC models, it also showed varying degrees of improvement compared with the other models, although these improvements were not pronounced. Specifically, the P values versus Rad-RF, DL-RF, DLR-RF, Clinic-LR, and RDC-RF were 0.116, 0.095, 0.173, 0.003, and 0.026, respectively (Supplementary Figure S2).

Table 3
Predictive performance of the individual and fusion models in the training and test cohorts.

Model		Accuracy	AUC	95% CI	Sensitivity	Specificity	PPV	NPV	Precision
Rad-RF	Train	0.840	0.906	0.8442–0.9688	0.911	0.750	0.820	0.871	0.820
	Test	0.829	0.812	0.6596–0.9635	0.913	0.667	0.840	0.800	0.840
DL-RF	Train	0.802	0.815	0.7190–0.9112	0.822	0.778	0.822	0.778	0.822
	Test	0.800	0.813	0.6684–0.9584	0.783	0.833	0.900	0.667	0.900
Clinic-LR	Train	0.630	0.611	0.5084–0.7138	0.778	0.444	0.636	0.615	0.636
	Test	0.743	0.665	0.5077–0.8220	0.913	0.417	0.750	0.714	0.750
RDC-RF	Train	0.852	0.948	0.9070–0.9893	0.822	0.889	0.902	0.800	0.902
	Test	0.800	0.850	0.7138–0.9854	0.739	0.917	0.944	0.647	0.944
DLR-RF	Train	0.852	0.931	0.8826–0.9798	0.956	0.722	0.811	0.929	0.811
	Test	0.829	0.855	0.7167–0.9935	0.783	0.917	0.947	0.687	0.947
RCLD-SVM	Train	0.926	0.965	0.9290–1.0000	0.956	0.889	0.915	0.941	0.915
	Test	0.914	0.924	0.8221–1.0000	0.913	0.917	0.955	0.846	0.955

AUC, area under the receiver operating characteristic curve; CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value; RF, random forest; LR, logistic regression; SVM, support vector machine; Rad, radiomics; DL, deep learning; Clinic, clinical; RDC, radiomics–clinical fusion; DLR, deep learning–radiomics fusion; RCLD, radiomics–clinical–deep learning fusion.

The RCLD model predicted OS risk stratification

The RCLD score was additionally employed to divide patients into high-risk and low-risk categories. The threshold established in the training cohort was carried over without modification to the test cohort. In the training cohort, 40 patients were classified as high risk and 41 as low risk. High-risk patients had significantly worse OS than low-risk patients (log-rank p = 0.003; Fig. 2g). In the test cohort, the identical locked cutoff assigned 17 patients to the high-risk category and 18 to the low-risk category. Once again, OS was significantly poorer in the high-risk group than in the low-risk group (log-rank p = 0.027; Fig. 2h). These results indicated that the RCLD score was associated with OS risk stratification in both cohorts.

Construction of the response probability nomogram

A pretreatment response probability nomogram was constructed using multivariable logistic regression by integrating liver cirrhosis, the Rad score and the DL score (Fig. 3a). Each variable was assigned an individual point value according to its regression coefficient. The points corresponding to the three predictors were summed to generate a total score, which was subsequently converted into the predicted probability of achieving an objective response. This nomogram was developed separately from the Clinic model and did not alter the definition of the Clinic model, which included liver cirrhosis alone.

(a) Predictive nomogram incorporating cirrhosis status, the Rad score, and the DL score for estimating the probability of TR. (b) SHAP summary plot of the Rad signature model, demonstrating the relevance of the four selected radiomic features and their contributions to the model predictions. (c) Representative Grad-CAM visualizations showing the image regions contributing to ResNet101 predictions in pretreatment arterial-phase CT images from the nTR and TR groups.

DL, deep learning; Grad-CAM, gradient-weighted class activation mapping;Rad, radiomics; SHAP, Shapley Additive exPlanations; TR, treatment responder; nTR,no- treatment responder.

Discussion

In this study, we developed a pretreatment CECT-based RCLD model to predict early objective response to HTIs in patients with HCC and PVTT. The model integrated cirrhosis status, four handcrafted radiomic features, and ResNet101-derived deep-learning features. This model achieved the highest area under the AUC point estimate among the six evaluated models, indicating that clinical, radiomic, and deep-learning information may provide complementary predictive value. However, in the internal test cohort, DeLong testing showed significant improvement only over the clinical-only and radiomics–deep-learning models; therefore, the numerical superiority over the other models should not be overstated. The RCLD-SVM score also stratified OS, with poorer survival in the high-risk group. This finding suggests that imaging patterns associated with early treatment resistance may also capture aggressive tumor biology, although the survival analysis should be considered exploratory rather than evidence of a causal prognostic effect.

These results are clinically relevant because no universally accepted optimal strategy exists for HCC with PVTT. European and US guidelines generally favor systemic, particularly immune-checkpoint inhibitor-based, therapy for macrovascular invasion[5], which is supported by the IMbrave150 and the early-phase KEYNOTE-524 studys[20,21]. In China, locoregional therapy is frequently combined with systemic treatment. Although TACE- and HAIC-based combinations have shown activity[22,23], TACE may not be appropriate when the main portal vein is totally occluded or when collateral circulation is insufficient.[24]. Because HAIC is nonembolizing, preserves arterial patency, and delivers high local drug concentrations, it may be valuable when TACE is contraindicated or ineffective, especially with main portal-vein invasion[9,25]. Nevertheless, biological heterogeneity means that some patients experience toxicity and financial burden without benefit[26], underscoring the need for pretreatment prediction. Previous models support imaging-based assessment: Hua et al. reported a validation AUC of approximately 0.89 for a CT-radiomics model[27], whereas Lu et al. reported training and test AUCs of 0.95 and 0.84 for an MRI-based clinical-radiomic model[28]. Our clinical-only model performed worst, with AUCs of 0.611 and 0.665, and cirrhosis was its only retained variable. Cirrhosis may summarize fibrosis, altered hepatic hemodynamics, impaired reserve, chronic inflammation, and immune dysfunction[29,30], but it should be regarded as host context rather than a tumor-specific or causal predictor. SHAP ranked gradient_firstorder_Minimum as the most influential radiomic feature, followed by wavelet_HLL_glrlm_GrayLevelVariance, wavelet_HHL_ngtdm_Busyness, and wavelet_HLL_ngtdm_Strength. Together, these features may describe local intensity transitions, gray-level dispersion, fine spatial variation, and coarse texture, potentially reflecting the distribution of viable tumor, necrosis, and stroma. The association of gradient_firstorder_Minimum with F12 expression and coagulation-related activity has been reported in gastric cancer [31]. In clinical feature screening, cirrhosis was retained as the sole predictive feature. Because cirrhosis frequently alters hepatic texture and gradient_firstorder_Minimum was both the radiomic feature with the greatest contribution in this study and

a corresponding texture feature, the clinical and radiomic features appeared mutually concordant.ResNet101 may additionally capture nonlinear morphology and spatial context[32]. Consistently, Grad-CAM showed relatively diffuse or peripheral attention, sometimes extending toward PVTT, in patients nTR, but more concentrated intratumoral hotspots in patients with TR.

Several limitations require emphasis. First, the retrospective, single-center design and small sample increased the risks of selection bias, unstable feature selection, overfitting, and optimistic performance estimates. A random split from one institution provides only internal validation and cannot demonstrate transportability across scanners, acquisition protocols, patient populations, or treatment practices. Although out-of-fold stacking reduced direct information leakage, it did not eliminate these risks. Future studies should evaluate a locked model in larger, temporally and geographically independent multicenter cohorts and report discrimination, calibration, and clinical utility. Second, 87.9% of patients had hepatitis B virus infection and most received lenvatinib, limiting applicability to other etiologies and systemic regimens. REFLECT subgroup findings in Greater China do not establish that hepatitis B virus infection predicts greater HTIs benefit. Third, HTIs is not a universally guideline-endorsed first-line standard for HCC with PVTT, and the model estimates response among treated patients rather than comparative benefit versus another therapy; therefore, it cannot yet select the optimal treatment. Fourth, OS stratification and Grad-CAM findings were exploratory, and the biological meaning of the selected features remains unverified. Prospective trials should assess clinical impact, compare treatment alternatives, standardize imaging and response assessments, and link imaging signatures with pathological, molecular, and conversion-surgery findings.

In conclusion, pretreatment CECT provides useful information for estimating early HTIs response in HCC with PVTT. By combining radiomic texture, ResNet101-derived imaging representations, and cirrhosis-related host context, the RCLD model showed the strongest overall discrimination and identified groups with different survival outcomes. Clinically, it may support noninvasive pretreatment risk stratification and individualized discussions of expected benefits, toxicity, and follow-up intensity.

This study has several limitations. First, its retrospective, single-center design and modest sample size (116 patients, including 35 in the test cohort) increase the risks of selection bias, model instability, and optimistic performance estimates. Although out-of-fold stacking reduced information leakage, a random split from the same source provides internal validation rather than evidence of transportability. Second, 87.9% of participants had HBV-related disease and most received lenvatinib, limiting generalizability across etiologies, populations, scanners, acquisition protocols, and systemic regimens. Third, tumor and PVTT segmentation was performed manually, while the deep-learning model used only the largest axial tumor slice; these choices may reduce reproducibility and incompletely capture three-dimensional heterogeneity. Fourth, early response assessed by mRECIST at 6–8 weeks is a surrogate endpoint and may not translate into durable survival benefit. The OS analysis was exploratory, its cutoff was training-derived, and the biological basis of the selected features remains uncertain. To address these limitations, we plan to prospectively enroll consecutive patients from multiple geographically independent centers. The preprocessing pipeline, model parameters, and decision threshold will be locked before external validation. Future studies will standardize imaging and response assessment, quantify interobserver segmentation reproducibility, and evaluate discrimination, calibration, Brier score, and decision-curve net benefit by center and clinically relevant subgroup. If calibration drift occurs, recalibration will be undertaken in a separate development dataset and independently retested. Larger cohorts with longer follow-up should evaluate progression-free and overall survival and determine whether model-guided decisions improve patient outcomes in a prospective clinical-impact study.

Conclusion

In this study, we developed and validated a pretreatment CECT-based multimodal fusion model integrating radiomics, deep learning, and liver cirrhosis for noninvasive prediction of early response to HTIs therapy in patients with HCC-PVTT. Among the fusion models, RCLD mobel demonstrated satisfactory performance in predicting HTIs treatment response (test AUC, 0.924; 95% CI, 0.822–1.000) and OS risk stratification. It could be used to identify potential candidates for HTIs therapy and guide clinical decision-making, although further studies are needed to confirm its clinical utility.

Abbreviations

AdaBoost
Adaptive boosting
AFP
Alpha-fetoprotein
ALBI
Albumin-bilirubin
ALT
Alanine aminotransferase
AST

Aspartate aminotransferase
AUC
Area under the receiver operating characteristic curve
BCLC
Barcelona Clinic Liver Cancer
BMI
Body mass index
BN
Batch normalization
CECT
Contrast-enhanced computed tomography
CI
Confidence interval
Clinic
Clinical model
Conv
Convolution
CR
Complete response
DCA
Decision curve analysis
DCR
Disease control rate
DL
Deep learning
DLR
Deep learning-radiomics fusion model
ECOG
Eastern Cooperative Oncology Group
F12
Coagulation factor XII gene
FC
Fully connected layer
GLRLM
Gray-level run length matrix
Grad-CAM
Gradient-weighted class activation mapping
HAIC
Hepatic arterial infusion chemotherapy
HBsAg
Hepatitis B surface antigen
HBV
Hepatitis B virus
HCC
Hepatocellular carcinoma
HHL
High-high-low wavelet subband
HLL
High-low-low wavelet subband
HTIs
Hepatic arterial infusion chemotherapy combined with tyrosine kinase inhibitors and PD-1 inhibitors
KM
Kaplan-Meier
KNN
K-nearest neighbors

LASSO
Least absolute shrinkage and selection operator
LightGBM
Light Gradient Boosting Machine
LR
Logistic regression
mRECIST
Modified Response Evaluation Criteria in Solid Tumors
MRI
Magnetic resonance imaging
NGTDM
Neighboring gray tone difference matrix
NPV
Negative predictive value
NSFC
National Natural Science Foundation of China
nTR
No treatment response
OR
Odds ratio
ORR
Objective response rate
OS
Overall survival
PCA
Principal component analysis
PD
Progressive disease
PD-1
Programmed cell death protein 1
PFS
Progression-free survival
PPV
Positive predictive value
PR
Partial response
PT
Prothrombin time
PVTT
Portal vein tumor thrombosis
Rad
Radiomics
RCLD
Radiomics-clinical-deep learning fusion model
RDC
Radiomics-clinical fusion model
RECIST
Response Evaluation Criteria in Solid Tumors
ReLU
Rectified linear unit
ResNet101
101-layer residual neural network
RF
Random forest
ROI

Region of interest
ROC
Receiver operating characteristic
SD
Stable disease (treatment response) / standard deviation (descriptive statistics)
SHAP
SHapley Additive exPlanations
SPNSF
Shandong Provincial Natural Science Foundation
SPSS
Statistical Package for the Social Sciences
SVM
Support vector machine
TACE
Transarterial chemoembolization
TBIL
Total bilirubin
TKI
Tyrosine kinase inhibitor
TNM
Tumor-node-metastasis
TR
Treatment response.

Declarations

Ethics approval and consent to participate

Following the Declaration of Helsinki, the study received ethical approval from the Ethics Committee of Shandong Cancer Hospital and Institute, Shandong First Medical University and Shandong Academy of Medical Sciences (Approval No.202607051). The requirement for informed consent was waived.

Competing interests

The authors declare no competing interests.

Funding

This work was supported by the National Natural Science Foundation of China (NSFC: 8227071910), the Shandong Provincial Natural Science Foundation (SPNSF: ZR2021MH060).

Author Contribution

GW: Conceptualization, Methodology, Data curation, Formal analysis, Visualization, Writing-original draft, Writing-review & editing. FL: Methodology, Data curation, Validation, Writing-review & editing. CW: Investigation, Data curation, Formal analysis, Writing-review & editing. BZ: Investigation, Data curation, Writing-review & editing. HZ: Investigation, Data curation, Writing-review & editing. HC: Investigation, Data curation, Visualization, Writing-review & editing. YS: Investigation, Data curation, Writing-review & editing. XR: Investigation, Validation, Writing-review & editing. Jiadong Liu: Investigation, Data curation, Writing-review & editing. JH: Conceptualization, Project administration, Supervision, Writing-review & editing.

Data Availability

All relevant data generated or analyzed during this study are included within the article and its supplementary information files. The raw datasets are available from the corresponding author upon reasonable request.

References

1. Sung, H. et al. Global cancer statistics 2024: GLOBOCAN estimates of incidence and mortality worldwide for 34 cancers in 186 countries. *Cancer J. Clin.* **76**, e70090. 10.3322/caac.70090 (2026).
2. Gilles, H., Garbutt, T., Landrum, J., Hepatocellular & Carcinoma *Crit. Care Nurs. Clin. N. Am.* **34**, 289–301, doi:10.1016/j.cnc.2022.04.004 (2022).
3. Tao, Z. W., Cheng, B. Q., Zhou, T. & Gao, Y. J. Management of hepatocellular carcinoma patients with portal vein tumor thrombosis: A narrative review. *Hepatobiliary Pancreat. Dis. international: HBPD INT.* **21**, 134–144. 10.1016/j.hbpd.2021.12.004 (2022).
4. Lu, J. et al. Management of patients with hepatocellular carcinoma and portal vein tumour thrombosis: comparing east and west. *lancet Gastroenterol. Hepatol.* **4**, 721–730. 10.1016/s2468-1253(19)30178-5 (2019).
5. EASL Clinical Practice Guidelines on the management. of hepatocellular carcinoma. *J. Hepatol.* **82**, 315–374. 10.1016/j.jhep.2024.08.028 (2025).
6. Singal, A. G. et al. AASLD Practice Guidance on prevention, diagnosis, and treatment of hepatocellular carcinoma. *Hepatol. (Baltimore Md)*. **78**, 1922–1965. 10.1097/hep.0000000000000466 (2023).
7. Reig, M. et al. BCLC strategy for prognosis prediction and treatment recommendations: The 2026 update. *J. Hepatol.* **84**, 631–654. 10.1016/j.jhep.2025.10.020 (2026).
8. Gavriilidis, P., Pawlik, T. M. & Azoulay, D. Comprehensive review of hepatocellular carcinoma with portal vein tumor thrombus: State of art and future perspectives. *Hepatobiliary Pancreat. Dis. international: HBPD INT.* **23**, 221–227. 10.1016/j.hbpd.2023.10.009 (2024).
9. Wu, C. et al. Hepatic arterial infusion chemotherapy combined with lenvatinib and programmed death receptor-1 inhibitors for hepatocellular carcinoma with Vp4 portal vein tumor thrombus: a multicenter, propensity score matching comparative study. *Hep. Intl.* **20**, 396–407. 10.1007/s12072-025-10884-6 (2026).
10. Zhang, T. Q. et al. Camrelizumab (a PD-1 inhibitor) plus apatinib (an VEGFR-2 inhibitor) and hepatic artery infusion chemotherapy for hepatocellular carcinoma in Barcelona Clinic Liver Cancer stage C (TRIPLET): a phase II study. *Signal. Transduct. Target. therapy*. **8**, 413. 10.1038/s41392-023-01663-6 (2023).
11. Yang, C. et al. Targeting the immune privilege of tumor-initiating cells to enhance cancer immunotherapy. *Cancer cell* 42, 2064–2081.e (2019). 10.1016/j.ccell.2024.10.008 (2024).
12. Gillies, R. J., Kinahan, P. E., Hricak, H. & Radiomics Images Are More than Pictures, They Are Data. *Radiology* **278**, 563–577. 10.1148/radiol.2015151169 (2016).
13. Lambin, P. et al. Radiomics: extracting more information from medical images using advanced feature analysis. *European journal of cancer (Oxford, England*:) 48, 441–446,) 48, 441–446, (1990). 10.1016/j.ejca.2011.11.036 (2012).
14. Park, H. J., Park, B. & Lee, S. S. Radiomics and Deep Learning: Hepatic Applications. *Korean J. Radiol.* **21**, 387–401. 10.3348/kjr.2019.0752 (2020).
15. Bai, W. et al. From Tumor to Tumor-Spleen: MRI Habitat Heterogeneity for Predicting Immunotherapy Outcome in Advanced Hepatocellular Carcinoma. *J. hepatocellular carcinoma*. **13**, 615844. 10.2147/jhc.S615844 (2026).
16. Lu, D. et al. Multimodal Modeling Distinguishes Treatment Response from Overall Survival in Hepatocellular Carcinoma Receiving Combined Interventional and Targeted Immunotherapy. *J. hepatocellular carcinoma*. **13**, 613014. 10.2147/jhc.S613014 (2026).
17. Bao, W. et al. Habitat and peritumoral CT radiomics accurately predict early treatment response to hepatic arterial infusion chemotherapy combined with tyrosine kinase inhibitors and programmed death-1 inhibitors in unresectable hepatocellular carcinoma. *Front. Oncol.* **16**, 1820483. 10.3389/fonc.2026.1820483 (2026).
18. Shi, G. Y. et al. 1D Gradient-Weighted Class Activation Mapping, Visualizing Decision Process of Convolutional Neural Network-Based Models in Spectroscopy Analysis. *Anal. Chem.* **95**, 9959–9966. 10.1021/acs.analchem.3c01101 (2023).
19. Molodianovitch, K., Faraggi, D. & Reiser, B. Comparing the areas under two correlated ROC curves: parametric and non-parametric approaches. *Biom. J.* **48**, 745–757. 10.1002/bimj.200610223 (2006).
20. Kudo, M. et al. IMbrave150: Efficacy and Safety of Atezolizumab plus Bevacizumab versus Sorafenib in Patients with Barcelona Clinic Liver Cancer Stage B Unresectable Hepatocellular Carcinoma: An Exploratory Analysis of the Phase III Study. *Liver cancer*. **12**, 238–250. 10.1159/000528272 (2023).
21. Finn, R. S. et al. Phase Ib Study of Lenvatinib Plus Pembrolizumab in Patients With Unresectable Hepatocellular Carcinoma. *J. Clin. oncology: official J. Am. Soc. Clin. Oncol.* **38**, 2960–2970. 10.1200/jco.20.00808 (2020).
22. Lin, Z. et al. Clinical efficacy of HAIC (FOLFOX) combined with lenvatinib plus PD-1 inhibitors vs. TACE combined with lenvatinib plus PD-1 inhibitors in the treatment of advanced hepatocellular carcinoma with portal vein tumor thrombus and arterioportal fistulas. *Am. J. cancer Res.* **13**, 5455–5465 (2023).

23. Ou, X. et al. Efficacy of Lenvatinib Combined with Anti-PD-1 Antibodies Plus Transcatheter Arterial Chemoembolization for Hepatocellular Carcinoma with Portal Vein Tumor Thrombus: A Retrospective, Multicenter Study. *Cancer Res. Treat.* **56**, 1207–1218. 10.4143/crt.2023.1165 (2024).
24. Xiang, X. et al. Transarterial chemoembolization versus best supportive care for patients with hepatocellular carcinoma with portal vein tumor thrombus: a multicenter study. *Eur. J. Surg. oncology: J. Eur. Soc. Surg. Oncol. Br. Association Surg. Oncol.* **45**, 1460–1467. 10.1016/j.ejso.2019.03.042 (2019).
25. Hu, J. et al. Hepatic Arterial Infusion Chemotherapy Using Oxaliplatin Plus 5-Fluorouracil Versus Transarterial Chemoembolization/Embolization for the Treatment of Advanced Hepatocellular Carcinoma with Major Portal Vein Tumor Thrombosis. *Cardiovasc. Interv. Radiol.* **43**, 996–1005. 10.1007/s00270-019-02406-3 (2020).
26. Li, G. et al. A habitat radiomics model based on contrast-enhanced MRI for predicting early treatment response to hepatic arterial infusion chemotherapy in patients with unresectable hepatocellular carcinoma. *Cell. Oncol. (Dordrecht Netherlands)*. **49** 10.1007/s13402-026-01196-4 (2026).
27. Hua, Y. et al. Pretreatment CT-based machine learning radiomics model predicts response in unresectable hepatocellular carcinoma treated with lenvatinib plus PD-1 inhibitors and interventional therapy. *J. Immunother. Cancer*. **12** 10.1136/jitc-2024-008953 (2024).
28. Lu, D. et al. MRI Radiomics to Predict Early Treatment Response to TACE Combined with Lenvatinib Plus a PD-1 Inhibitor for Hepatocellular Carcinoma with Portal Vein Tumor Thrombus. *J. hepatocellular carcinoma*. **12**, 985–998. 10.2147/jhc.S513696 (2025).
29. Schuppan, D. & Afdhal, N. H. Liver cirrhosis. *Lancet (London England)*. **371**, 838–851. 10.1016/s0140-6736(08)60383-9 (2008).
30. Albillos, A., Lario, M. & Álvarez-Mon, M. Cirrhosis-associated immune dysfunction: distinctive features and clinical relevance. *J. Hepatol.* **61**, 1385–1396. 10.1016/j.jhep.2014.08.010 (2014).
31. Pan, S. et al. Radiomic model for preoperative prediction of mismatch repair deficiency in gastric cancer: a multicenter study integrating tumor sub-region radiomics and transcriptomics. *Cancer biology Med.* **22**, 1210–1217. 10.20892/j.issn.2095-3941.2025.0254 (2025).
32. He, J. et al. Development of a deep learning model for T1N0 gastric cancer diagnosis using 2.5D radiomic data in preoperative CT images. *NPJ precision Oncol.* **9**, 249. 10.1038/s41698-025-01055-9 (2025).

Figures

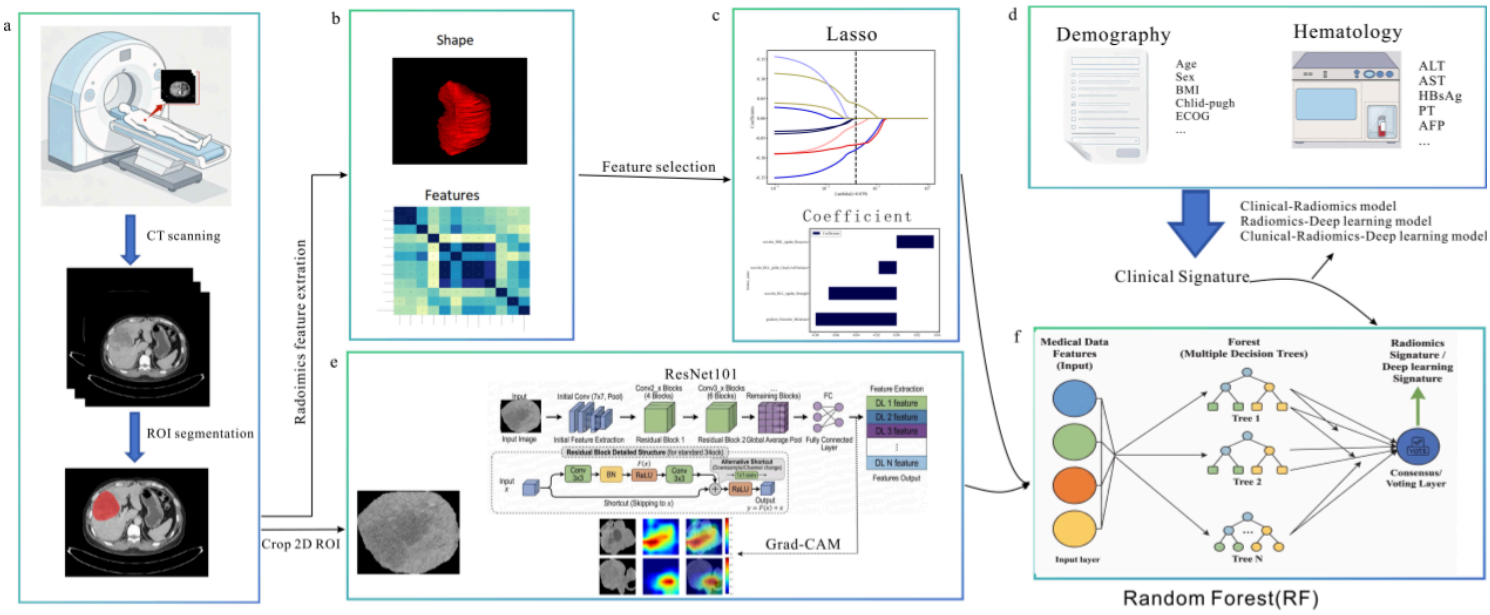


Figure 1

Workflow of the clinical–radiomics–deep learning model development.

(a)CT image acquisition and segmentation. (b) Radiomics feature extraction. (c) Radiomics features

selection. (d) Deep learning feature extraction and visualization. (e) Clinical feature selection and multimodal integration. (f) Classifier comparison and final model construction.

Grad-CAM, gradient-weighted class activation mapping; RF, random forest; ROI, region of interest.

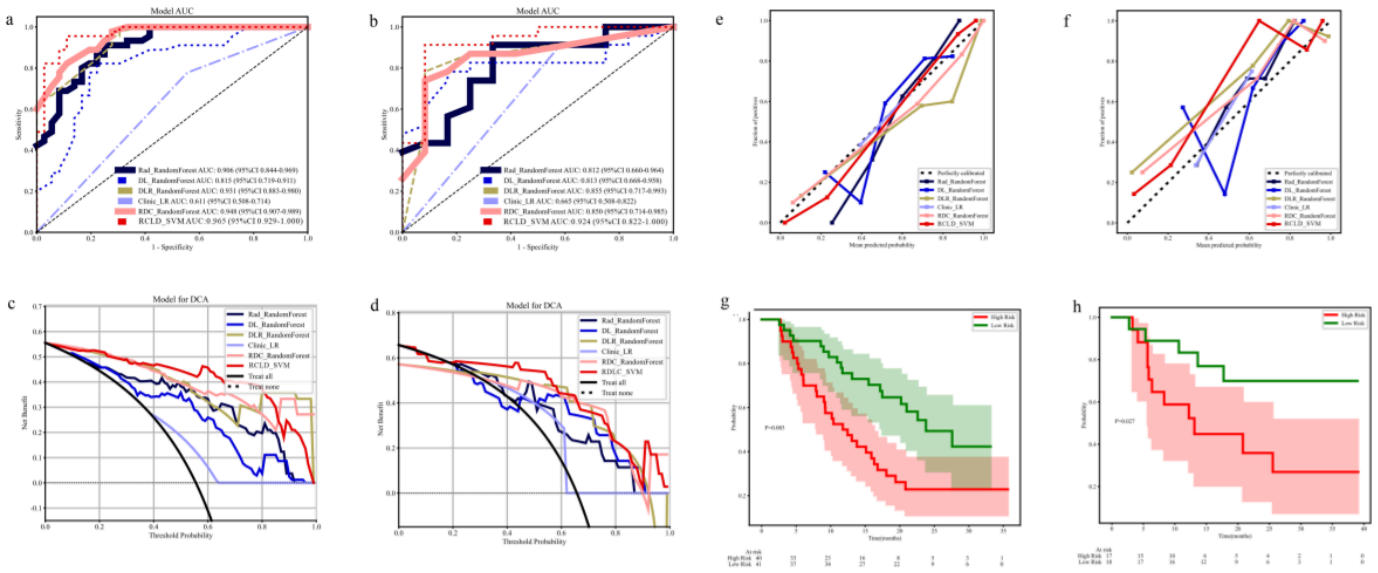


Figure 2

Predictive performance, calibration, clinical utility, and OS stratification of the developed models.

(a, b) ROC curves in the training (a) and test (b) cohorts, with the RCLD model achieving the highest AUC. (c, d) Decision curve analyses of the training (c) and test (d) cohorts. (e, f) Calibration curves showing agreement between predicted and observed probabilities. (g, h) Kaplan–Meier curves stratified by the RCLD score, showing poorer OS in the high-risk group.

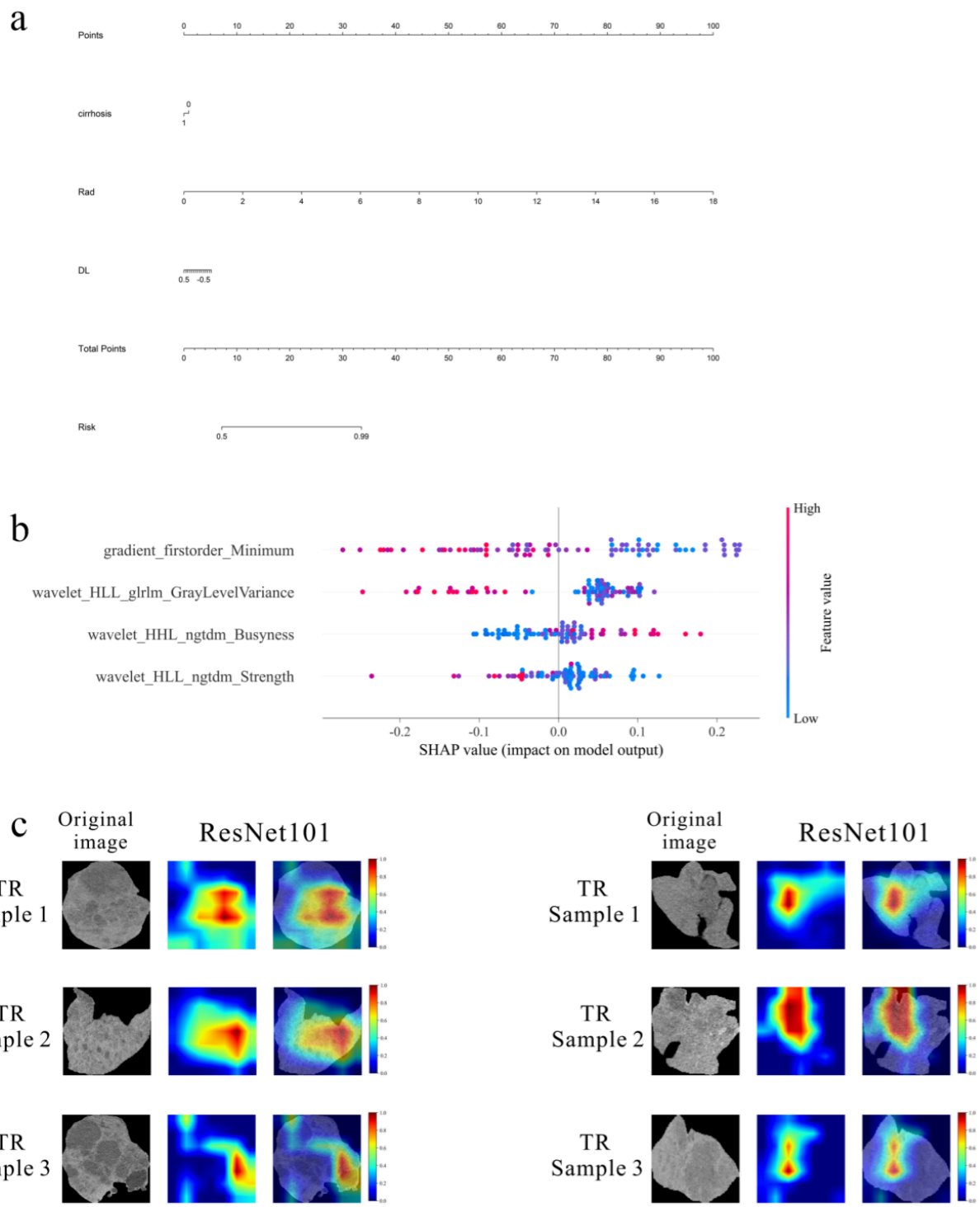


Figure 3

Nomogram and interpretability analyses of the multimodal prediction model.

(a) Predictive nomogram incorporating cirrhosis status, the Rad score, and the DL score for estimating the probability of TR. (b) SHAP summary plot of the Rad signature model, demonstrating the relevance of the four selected radiomic features and their contributions to the model predictions. (c) Representative Grad-CAM visualizations showing the image regions contributing to ResNet101 predictions in pretreatment arterial-phase CT images from the nTR and TR groups.

DL, deep learning; Grad-CAM, gradient-weighted class activation mapping;Rad, radiomics; SHAP, Shapley Additive exPlanations; TR, treatment responder; nTR,no- treatment responder.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplement.pdf