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

Electronic Supplementary Material

1. Supplementary Methods

Appendix E1 Data preparation and patient selection

Patients were included if they met the following criteria: (a) age of 18–80 years; (b) clinically or histologically confirmed HCC, with PVTT identified on contrast-enhanced CT or magnetic resonance imaging; (c) Child-Pugh class A or B and an Eastern Cooperative Oncology Group performance status (ECOG-PS) score of 0 or 1; (d) completion of at least one cycle of HTI, followed by an initial imaging assessment; (e) pretreatment CECT images suitable for radiomics analysis; and (f) at least one measurable intrahepatic lesion according to the modified Response Evaluation Criteria in Solid Tumors (mRECIST).Patients were excluded if they had another primary malignancy, Child-Pugh class C liver function, or severe cardiac, pulmonary, or renal disease. Those with incomplete clinical records or without the pretreatment or follow-up imaging required for radiomics analysis and response assessment were also excluded. Finally, a total of 116 eligible patients were enrolled. Subsequently, these patients were randomly assigned to two different cohorts at a

ratio of 7:3 using stratified randomization to preserve the proportion of treatment response. The training cohort comprised 81 patients, while the test cohort included 35 patients.

Demographic and clinical data encompassed age, sex, and body mass index, together with hepatitis B or C virus infection status and cirrhosis. Performance status and liver function were assessed using the ECOG-PS score and Child-Pugh class, respectively, while tumor burden was characterized by Barcelona Clinic Liver Cancer (BCLC) stage, tumor number, maximum tumor diameter, PVTT classification according to the Vp classification system, and extrahepatic metastasis status. Laboratory variables comprised alpha-fetoprotein level and routine hematological and biochemical indices. Radiomics analysis was based on pretreatment CECT images stored in Digital Imaging and Communications in Medicine (DICOM) format.

Appendix E2 Implementation of the treatment and response assessment

E2.1 Administration of molecular-targeted agents and PD-1 inhibitors

The treatment regimen was determined by a multidisciplinary team based on the patients' physical condition, liver function, tumor burden, and treatment preferences and in consultation with the patients. All participants received oral molecular-targeted agents and intravenous PD-1 inhibitors during HAIC. First-line TKIs, including sorafenib or lenvatinib (Lenvima; Eisai Co., Ltd., Tokyo, Japan), were generally initiated 1–5 days after the first HAIC session and continued thereafter. Oral lenvatinib was administered at an initial dose of 12 mg once daily for patients

weighing ≥60 kg with Child-Pugh class A liver function or 8 mg once daily for those weighing <60 kg. The initial dose and subsequent dose adjustments were determined according to body weight, liver function, toxicity profile, and clinical status.

PD-1 inhibitor-based immunotherapy was generally initiated 1–3 days after HAIC and administered intravenously every 3 weeks at the recommended fixed dose. Treatment was continued until disease progression, intolerable adverse events (AEs), successful downstaging for curative resection, technical infeasibility, patient withdrawal, or death. The dosage of TKIs was reduced or temporarily interrupted when clinically significant treatment-related AEs occurred. In patients with radiological progression or intolerance to first-line TKIs, second-line agents, including regorafenib or apatinib, could be administered after multidisciplinary assessment. The severity of treatment-related AEs was evaluated according to the Common Terminology Criteria for Adverse Events version 5.0.

E2.2 Implementation of hepatic arterial infusion chemotherapy

The HAIC treatment schedule was determined by the multidisciplinary team according to the patients' physical condition, liver function, tumor characteristics, vascular anatomy, and treatment preferences. All procedures were performed by experienced interventional radiologists according to standardized protocols described in previous studies[1,2].

The HAIC procedure was performed by puncturing the femoral artery using the modified Seldinger technique. Digital subtraction angiography was subsequently performed using a Philips FD20 angiography system (Philips Healthcare, Amsterdam, the Netherlands). A 5-Fr Yashiro catheter (Terumo, Tokyo, Japan) was introduced into the celiac trunk and superior mesenteric artery for angiography to identify the tumor-feeding hepatic arteries. Subsequently, a 2.7-Fr microcatheter (Terumo, Tokyo, Japan) was super-selectively advanced into the tumor-feeding arteries whenever technically feasible. The catheter tip was retained within the target artery, and the external catheter segment was securely fixed in the inguinal region during chemotherapy infusion.A modified FOLFOX6 regimen was adopted for HAIC. On day 1, oxaliplatin was administered at a dose of 130 mg/m² over 3 h, followed by leucovorin at 200 mg/m² over 3–5 h and fluorouracil at 400 mg/m² as an intravenous bolus. This was followed by continuous infusion of fluorouracil at a dose of 2,400 mg/m² over 23–46 h. Treatment cycles were repeated every 3 weeks, and patients generally received 4–6 cycles unless intrahepatic disease progression, technical infeasibility, or unacceptable toxicity occurred.

E2.3 Response assessment

All enrolled patients received at least one cycle of HTI and completed the first post-treatment imaging evaluation. According to the mRECIST criteria, two abdominal radiologists with more than 10 years of clinical experience independently evaluated the treatment response by comparing the target lesions on pretreatment and

post-treatment imaging examinations, including CECT and contrast-enhanced MRI. The first post-treatment imaging evaluation was performed 6–8 weeks after treatment initiation. Any disagreement between the two radiologists was resolved by consensus with a third senior abdominal radiologist with more than 20 years of clinical experience. Treatment response was assigned to one of four categories—complete response (CR), partial response (PR), stable disease (SD), and progressive disease (PD). For CR, arterial-phase enhancement had to disappear in all target lesions; for PR, the sum of viable target lesion diameters needed to fall by 30% or more from baseline. PD was defined as an increase of at least 20% in the sum of the diameters of viable target lesions, using the smallest recorded sum as the reference, or the appearance of new lesions. SD was defined as neither sufficient tumor shrinkage to qualify as PR nor sufficient tumor enlargement to qualify as PD.

The primary outcome of this investigation was the early objective response at the first imaging evaluation after treatment. Participants who showed CR or PR were classified into the treatment response (TR) group, whereas those with SD or PD were classified into the no treatment response (nTR) group. ORR was calculated as the percentage of patients achieving CR or PR, whereas DCR represented the percentage achieving CR, PR, or SD.

Appendix E3 CECT Image Acquisition and Preprocessing

All patients underwent multiphasic CECT examination within 1 month before treatment using multidetector CT scanners manufactured by GE Healthcare or

Siemens Healthineers. CT acquisition settings were as follows: 120 kV tube voltage, automated tube-current modulation, 1.25-mm slice thickness (range, 0.75–1.50 mm), and a 512 × 512 image matrix.Contrast medium was administered intravenously at a dose of 1.5 mL/kg body weight and an injection rate of 3.0–3.5 mL/s. Contrast-enhanced imaging included arterial-phase acquisition at 25–30 s after contrast administration and portal venous-phase acquisition at 70–80 s. All CECT images were exported in DICOM format for subsequent tumor segmentation and radiomics analysis.

The pretreatment CECT images in DICOM format were retrieved from the picture archiving and communication system and de-identified before analysis. Raw CT images were subsequently converted from DICOM to Neuroimaging Informatics Technology Initiative (NIfTI) format using Python version 3.7. To reduce differences in spatial resolution among images acquired using different CT scanners, all images were resampled to an isotropic voxel spacing of 1 × 1 × 1 mm^3 using B-spline interpolation, while nearest-neighbor interpolation was applied to the corresponding segmentation masks. CT attenuation values were clipped to a uniform range of −1024 to 1024 Hounsfield units (HU). For radiomics and deep learning analysis, CT images were windowed using a window width of 350 HU and a window level of 50 HU and subsequently linearly normalized to a range of 0–1. All images were reviewed by two radiologists to ensure adequate image quality before analysis. All preprocessing parameters were defined using the training cohort and subsequently applied to the internal test cohort without modification.

Appendix E4 Region of interest segmentation and feature extraction

The primary HCC lesion and the associated portal vein tumor thrombus were manually delineated on CECT images by two abdominal radiologists with 8 and 12 years of clinical experience. Both radiologists were blinded to the histopathological findings. The contours covered the entire tumor and tumor thrombus but excluded the surrounding normal liver parenchyma, visible bile ducts, major non-tumorous intrahepatic vessels, and obvious imaging artifacts. Three-dimensional regions of interest (ROIs) were then generated by merging the contours drawn on consecutive slices. All ROIs were subsequently reviewed by a senior abdominal radiologist (Professor C, with more than 20 years of experience), who corrected inaccurate or questionable contours when necessary.

The radiomics extension implemented in PyRadiomics version 3.0.1 was used to extract quantitative radiomics features from the three-dimensional ROIs. For each ROI, 1,563 features were initially computed. These comprised shape and first-order statistical features, along with texture features obtained from the gray-level co-occurrence matrix, gray-level run-length matrix, gray-level size-zone matrix, neighboring gray-tone difference matrix, and gray-level dependence matrix. Feature calculation was performed on the original images and on filtered images produced by Laplacian of Gaussian, gradient, square, square-root, exponential, wavelet, and three-dimensional local binary pattern transformations. The definitions of features and

the settings used for their extraction conformed to Image Biomarker Standardization Initiative guidance when applicable.

Appendix E5 Radiomics, clinical and DL feature selection and model construction

To prevent overfitting of the developed radiomics model, all feature selection procedures were performed exclusively within the training cohort. The methods used to reduce the feature dimensionality were as follows. First, the extracted radiomics features were standardized through Z-score normalization using the means and standard deviations calculated from the training cohort, followed by the removal of low-variance features. An independent-samples t-test was subsequently performed to retain features showing significant differences between the TR and nTR groups ($P<0.05$). Second, Spearman correlation analysis was employed to calculate the pairwise correlations among the retained features. For each pair of features with an absolute correlation coefficient greater than 0.90, one redundant feature was excluded. In the final step, least absolute shrinkage and selection operator logistic (LASSO) regression combined with 10-fold cross-validation was applied to identify the optimal regularization parameter and to select the features used for constructing the radiomics model.

Clinical feature selection was also performed exclusively within the training cohort. For clinical feature screening, normally distributed continuous variables were expressed as the mean ± standard deviation and compared between groups using the

independent-samples t-test. Skewed continuous variables were presented as the median and interquartile range and compared using the Mann–Whitney U test. 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. Liver cirrhosis was the only variable meeting the prespecified statistical screening criterion and was therefore retained as the sole predictor for construction of the clinical model. Because no additional covariates entered the final clinical model, the reported cirrhosis estimate represents a one-predictor association rather than a fully confounder-adjusted independent effect. The selected predictor was applied unchanged to the test cohort, which was not involved in clinical feature selection or model development.

For each patient, the axial CT slice containing the largest cross-sectional area of the tumor within the ROI was selected for deep learning analysis. The CT window width and window level were set to 350 and 50 Hounsfield units, respectively. Only the ROI containing the tumor was retained, and the surrounding background information was excluded. Subsequently, the tumor region was cropped, resized to 224 × 224 pixels, and linearly normalized to a range of 0–1 before being entered into the deep learning network.To identify the optimal convolutional neural network for DL feature extraction, ResNet50, ResNet101, and DenseNet121 were compared according to their cross-validation performance in the training cohort. Among these networks, ResNet101 showed the best predictive performance and was therefore selected for subsequent analysis. To reduce the computational burden associated with training a

complex deep learning network and improve model convergence, transfer learning was performed using weights pretrained on the ImageNet dataset. The original classification layer of ResNet101 was replaced with a binary output layer. The network was trained using the stochastic gradient descent optimizer with an initial learning rate of 0.01 and cross-entropy loss. Model training was conducted for 40 epochs with a batch size of 32After network training, the final classification layer was removed, and the output of the global average pooling layer was used as the DL feature representation. Ultimately, the ResNet101 network generated 2,048 DL features from each CT image. Principal component analysis (PCA) was subsequently fitted in the training cohort to reduce these features to seven principal components. The seven principal components were then screened within the training cohort, and the three features most relevant to treatment response were retained for construction of the final DL model using the random forest algorithm. The predicted probability of objective response for each patient was defined as the DL signature. Network comparison, PCA fitting, DL feature screening, and random forest training were performed exclusively within the training cohort. The fitted feature transformations, selected features, and trained model were subsequently applied to the test cohort without further modification.

Appendix E6 Fusion model construction and classifier selection

Three multimodal fusion models were constructed using probability-level late fusion. The DLR model was generated by late fusion of the Rad and DL component models;

the RDC model was generated by late fusion of the Rad component model and the Clinic component model, in which liver cirrhosis was the sole clinical predictor; and the Rad-Clinic-DL(RCLD) model was generated by late fusion of only the Rad, DL, and Clinic component models. The meta-features for the DLR, RDC, and RCLD models were the predicted response probabilities from their corresponding component models. For all three models, SVM, RF, and LightGBM were used as candidate machine-learning models, and their performance was evaluated by stratified K-fold cross-validation within the training cohort using AUC, accuracy, sensitivity, specificity, and F1-score. To prevent stacking leakage, the training meta-features were generated from stratified K-fold out-of-fold (OOF) predictions. Within each fold, preprocessing, feature selection, and component-model fitting were repeated, and predictions were obtained only for the held-out samples. The aggregated OOF probabilities were then used for meta-learner training. Subsequently, all component models were refitted using the full training cohort, their probabilities were generated for the frozen test cohort, and the final predictions were produced using the corresponding trained meta-learners. All model evaluation and hyperparameter tuning were performed exclusively within the training cohort.

Appendix E7 Response Risk Nomogram and survival analysis

To support individualized pretreatment assessment, a response nomogram was constructed using multivariable logistic regression incorporating cirrhosis status, the Rad score, and the DL score. The two imaging scores represented quantitative

phenotypes of the primary tumor and PVTT, whereas cirrhosis provided readily available host-liver context. The nomogram was developed separately from the clinical-only model, which included cirrhosis as its sole predictor.

For prognostic assessment in patients with HCC and PVTT receiving HTI, the score from the final RCLD mobel was used to stratify OS. The cutoff maximizing the Youden index for objective response was determined in the training cohort and applied unchanged to the test cohort without further optimization. Patients were assigned to high- or low-risk groups according to the model-output direction, with the less favorable category designated high risk. OS was summarized using Kaplan–Meier curves, and differences were assessed using the log-rank test.

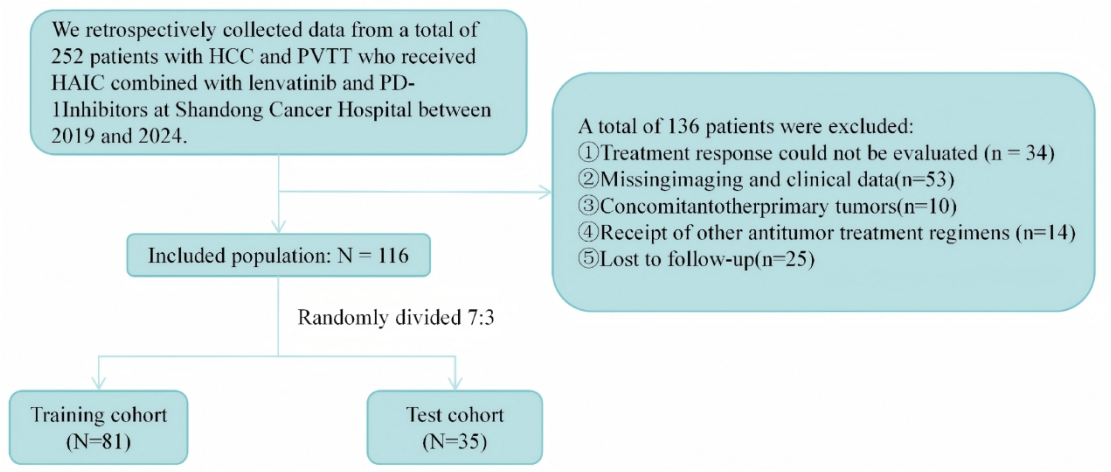
References

1 Wu, C. *et al.* TACE-HAIC versus HAIC combined with TKIs and ICIs for hepatocellular carcinoma with a high tumor burden-a propensity-score matching comparative study. *Frontiers in immunology* **16**, 1664756, doi:10.3389/fimmu.2025.1664756 (2025).

2 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. *Hepatology international* **20**, 396-407, doi:10.1007/s12072-025-10884-6 (2026).

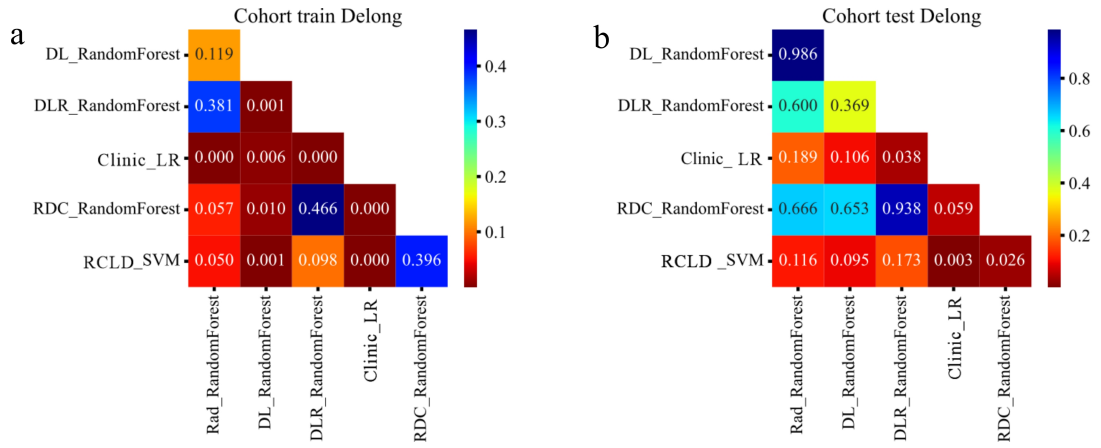
2. Supplementary Figures

Figure S1.Patient selection and cohort assignment flowchart.



Flowchart of the patient selection and cohort allocation. HCC: hepatocellular carcinoma; PVTT: portal vein tumor thrombosis; HAIC: hepatic arterial infusion chemotherapy; PD-1: programmed cell death protein 1.

Figure S2.DeLong test p-values for model comparisons.



Pairwise comparisons of the six models using DeLong's test in the training (a) and test (b) cohorts. Cell values represent P values; $P < 0.05$ indicates a statistically significant difference.

3. Supplementary Table

Table S1 CT protocols

Parameters		
CT version	Discovery (CT750 HD GE Healthcare, USA)	Somatom Force (Siemens Healthineers, Germany)
CT tube voltage	120 kV	120kv
CT tube current	50-350 mA	120 mA
CT rotation time	0.5 s	0.25 s
CT detector collimation	64 × 0.625 mm	128 × 0.6 mm
section thickness	5 mm	5mm
section interval	5 mm	5mm
Image matrix	512 × 512	512 × 512
Contrast agent type	Omnipaque	Ultravist
Contrast agent concentration	300mgI/mL	350mgI/mL
Contrast agent dosage	infused 1.5ml/kg body weigh	infused 1.2ml/kg body weigh
Contrast agent infused rate	3.5mL/s	3.5mL/s
Portal Venous Phase interval time	60-80s after the contrast injection	60-80s after the contrast injection
Reconstruction image thickness for Arterial phase	1.0mm	1.25mm