The Oncologist, 2025, **30**, oyaf013
https://doi.org/10.1093/oncolo/oyaf013
Advance access publication 20 March 2025
Original Article

Predicting early recurrence of hepatocellular carcinoma after thermal ablation based on longitudinal MRI with a deep learning approach

Qingyang Kong and Kai Li*

Department of Ultrasound, The Third Affiliated Hospital of Sun Yat-Sen University, Guangzhou, People's Republic of China

*Corresponding author: Kai Li, PhD, Department of Ultrasound, The Third Affiliated Hospital of Sun Yat-Sen University, Guangzhou, People's Republic of China. E-mail: likai@mail.sysu.edu.cn

Abstract

Background: Accurate prediction of early recurrence (ER) is essential to improve the prognosis of patients with hepatocellular carcinoma (HCC) underwent thermal ablation (TA). Therefore, a deep learning model system using longitudinal magnetic resonance imaging (MRI) was developed to predict ER of patients with HCC.

Methods: From 2014, April to 2017, May, a total of 289 eligible patients with HCC underwent TA were retrospectively enrolled from 3 hospitals and assigned into one training cohort (n = 254) and one external testing cohort (n = 35). Two deep learning models (Pre and PrePost) were developed using the pre-operative MRI and longitudinal MRI (pre- and post-operative) to predict ER for the patients with HCC after TA, respectively. Then, an integrated model (DL_Clinical) incorporating PrePost model signature and clinical variables was built for post-ablation ER risk stratification for the patients with HCC.

Results: In the external testing cohort, the area under the receiver operating characteristic curve (AUC) of the DL_Clinical model was better than that of the Clinical (0.740 vs 0.571), Pre (0.740 vs 0.648), and PrePost model (0.740 vs 0.689). Additionally, there was a significant difference in RFS between the high- and low-risk groups which were divided by the DL_Clinical model (P = .04).

Conclusions: The PrePost model developed using longitudinal MRI showed outstanding performance for predicting post-ablation ER of HCC. The DL_Clinical model could stratify the patients into high- and low-risk groups, which may help physicians in treatment and surveillance strategy selection in clinical practice.

Key words: hepatocellular carcinoma; thermal ablation; longitudinal MRI; deep learning; early recurrence.

Implications for practice

In the present study, several deep learning models (Pre and PrePost) were developed using longitudinal MRI to predict ER for patients with hepatocellular carcinoma (HCC) who underwent thermal ablation (TA), and the integrated model (DL_Clinical) which incorporated the PrePost signature and clinical variables yielded the best ER predictive performance in the external testing cohort, which may help doctors in surveillance strategy selection for patients with HCC who underwent TA.

Hepatocellular carcinoma (HCC) is the second leading cause of cancer-specific mortality globally, and more than half of these cases occurred in Asian-Pacific countries, especially in China.[1-3] Early intervention plays an important role in reducing the HCC related death and prolonging the lifespan for these patients.[2] Surgical resection (SR), liver transplantation (LT), and thermal ablation (TA) are recommended by multiple international guidelines practice as the first-line treatment in patients with HCC in early stage (Barcelona Clinic Liver Cancer [BCLC] A).[4,5] Among them, TA has many advantages including less trauma, fewer complication, and cost-effectiveness, and can be applied not only in patients with HCC ineligible for surgery as an alternative treatment but also to build a bridge for LT during the waiting process for liver source.[6-8]

Unfortunately, the 5-years recurrence rate with approximate 50% or more after TA remains a great challenge in clinical practice, greatly affecting the patients' with HCC prognosis.[9,10] Therefore, accurate prediction of early recurrence (ER) for patients with HCC is important for personalized surveillance strategy. Several previous studies focused on predicting HCC recurrence after SR or LT using clinical variables, imaging data or gene expression data, showing considerable predictive ability.[11-13] However, reliable tools for recurrence prediction in patients with HCC after TA remain limited and lack of multi-institutional validation.[14,15]

Deep learning (DL) based model, as emerging promising imaging analysis approaches, have been widely investigated in tumor characteristics and shown great value in various

Received: 23 October 2024; Accepted: 13 January 2025.

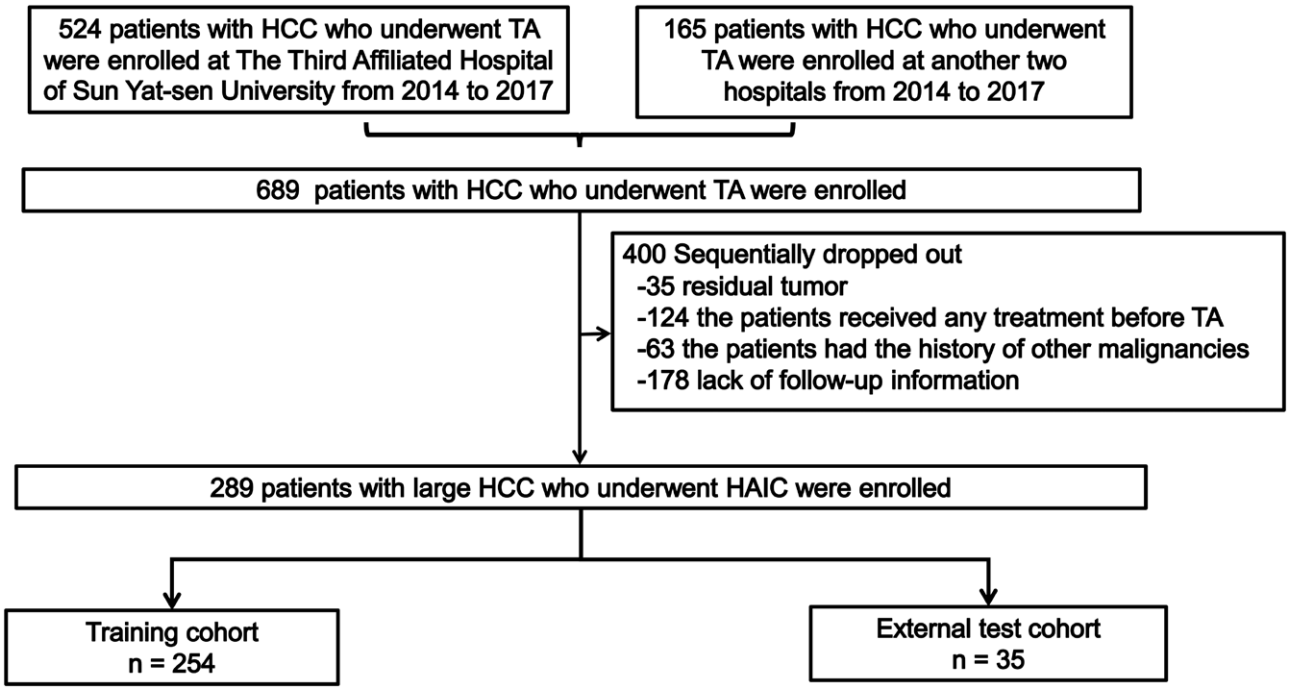

Figure 1. The enrollment pathway of patient with HCC who underwent thermal ablation. Abbreviations: HAIC, hepatic arterial infusion chemotherapy, HCC, hepatocellular carcinoma.

prediction tasks.[16,17] By using the DL method, abundant information from radiographic images can be extracted through a data-driven way. However, most ER prediction models are only based on pre-operative images, the performance of which are inherently limited because the abundant physiological information reflected by the ablation zones in the post-operative magnetic resonance imaging (MRI) are not taken into consideration.

Therefore, we aim to propose a reliable DL model for ER prediction of patients with HCC based on longitudinal MRI (ie, pre- and post-operative MRI) in this study, which is validated by an external testing cohort to provide guidance for treatment and surveillance strategies selection for patients with HCC who underwent TA.

Methods

Patients enrolled

This retrospective study complied with the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) guideline. This retrospective, multi-center study protocol was approved by the Institutional Review Board of Third Affiliated Hospital of Sun Yat-sen University following the principles of the 1975 Helsinki Declaration, and the written informed consent was waived because of the retrospective nature of this study.

From 2014, April to 2017, May, a total of 689 patients with early-stage HCC who underwent initial TA including microwave ablation and radiofrequency ablation at 3 high-volume medical centers (namely the Third Affiliated Hospital of Sun Yat-sen University [TAHSYU], the Sun Yat-sen University Cancer Center [SYUCC], and the Province Hospital of Fujian Medical University [PHFMU]) were reviewed. HCC was diagnosed according to the American Association for the Study of Liver Diseases (AASLD), and European Association for the Study of Liver (EASL).[4,5] The inclusion criteria were listed as following: (1) Eastern Cooperative Oncology Group performance scores < 2 scores: (2) Child-Turcotte-Pugh grade A; (3) maximum diameter of single tumor <5 cm or 2-3 tumors <3 cm; (4 absence of macrovascular invasion or metastasis; (5) enhanced MRI scanning within 4 weeks before TA. The exclusion criteria were as follows: (1) residual tumor; (2) history of other malignancies; (3) those who underwent other treatments such as surgery, transarterial chemoembolization [TACE], and chemoradiation before TA; (4) lack of follow-up information. Figure 1 demonstrates the patient enrollment pathway. Eligible patients were divided into the training cohort (the data were collected from TAHSYU) and testing cohort (another 2 hospitals).

Follow-up and recurrence assessment

The follow-up duration was terminated in September 2023. The serum AFP and contrast-enhanced imaging were examined again at 1 month after initial TA and at approximately 3- to 6-month intervals thereafter. The definition of ER was the lesions presenting abnormal nodular, disseminated, and/or unusual patterns of peripheral enhancement at the arterial phase and the washout at the delayed phase in the contrast-enhanced imaging examination, which were away from or abutting the ablated area within 2 years.[18] A typical example of the lesion was shown in **Figure 2**. Recurrence-free survival (RFS) was calculated by the period between the first session of TA and identification of recurrence or the last follow up, censoring recurrence-free patients at the date of last follow-up and those who died of other causes.

MR image review

The MRI data were taken from the picture archiving and communication system (PACS) database. The MRI data including 4 phases of T1WI including plain scan (PS),

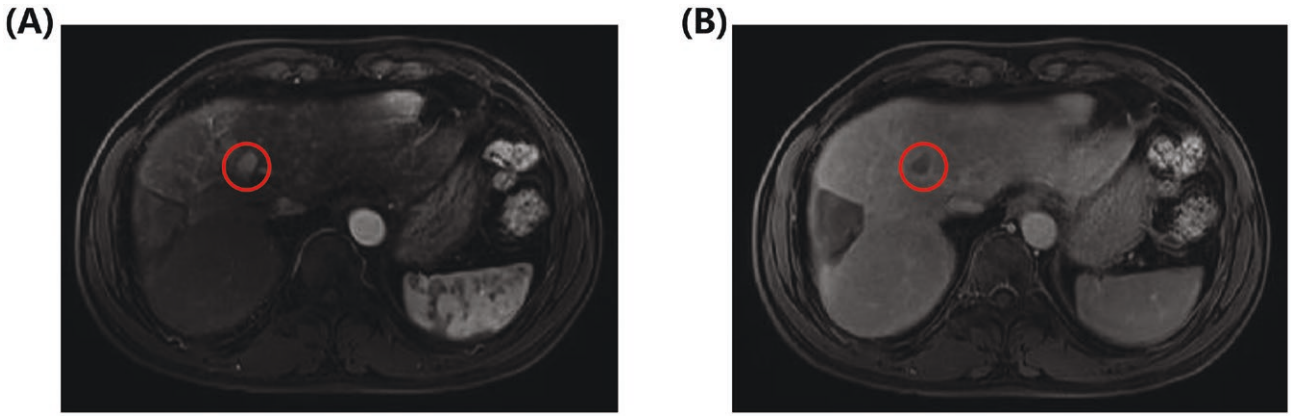

Figure 2. A 52 years old, male, patient with HCC who underwent MWA was reviewed. A high signal lesion with 15 mm which was away for the ablation zone was found at 23 months after MWA. (A) The lesion at the arterial phase. (B) The lesion at the delayed phase. Abbreviations: HCC, hepatocellular carcinoma; MWA, microwave ablation.

arterial portal (AP), portal vein phase (PVP), and delayed phase (DP). The images collected from the MRI examination before TA (ie, pre-operative) and the MRI examination at 1 month after TA (ie, post-operative) were used in the present study. The acquisition of T1WI parameters were shown in Supplementary Table S1. One radiologist (K.L., with 15 years of clinical experience) delineated the regions of interest (ROIs) of the tumor on the slice including the largest tumor profile in pre-operative DP, and the ablation zone on the slice including the largest ablation zone profile in post-operative DP. A secondary radiologist (C.A., with 10 years of clinical experience) reviewed and adjusted the ROIs. The delineation was implemented using the open-source ITK-SNAP software (http://www.itksnap.org/pmwiki/pmwiki.php).

Data preprocessing

The purpose of data preprocessing is to transform the raw data into a suitable format for DL models. In the present study, data preprocessing was implemented on the PS, AP, PVP, and DP of the pre- and post-operative T1WIs. There were 5 data preprocessing steps, including data de-identification, image registration, slices extraction, regions cropping, and data augmentation. Finally, 3 slices were extracted from each phase of T1WI, which included the slice with the largest profile of ROI, the slices above and below the one with the largest profile of ROI. Rectangular regions were cropped from the extracted slices, and the size of each of the cropped regions was 128 × 128 pixels for containing the peritumoral tissue microenvironment. The DL models were developed using the cropped regions. The flowchart of data preprocessing was shown in the solid border in Figure 3(A). The details of the data preprocessing steps were described in Supplementary Method 1.1.

Models construction

A total number of 289 patients with HCC underwent TA were used to develop and validate the models. Among them, 254 patients from the TAHSYU were used as the training cohort, of which 20% of the patients were randomly selected to compose a tuning cohort. The external testing cohort was consisted of 20 patients from the SYUCC and 15 patients from the PHFMU. Thus, the performance of the models can be validated on a completely independent external testing cohort. To explore and select the optimal model to predict ER for the patients with HCC after TA, 4 models were developed, including the Pre, PrePost, Clinical, and DL_Clinical models.

The flowchart of model construction is shown in Figure 3(B). First, 2 DL models were built with various input images (ie, PS, AP, PVP, and DP of T1WI). Specifically, one deep learning model was constructed to predict ER using the PS, AP, PVP, and DP of the pre-operative T1WI, which was referred to as the pre model. Furthermore, to explore the incremental recurrence predictive value of the post-operative T1WI, another DL model was constructed to predict ER using the PS, AP, PVP, and DP of both pre- and post-operative T1WIs, which was referred to as the PrePost model. Then, to investigate the value of independent clinical variables to the prediction of ER, the clinical variables of tumor size,[19] number of tumors,[19] HBV (viral hepatitis type B),[20] TP (total protein),[21] and AFP (α-fetoprotein)[19] were used to construct the Clinical model using the extra trees algorithm. Finally, an integrated model, namely the DL_Clinical model, was developed for investigating the improvements on the Clinical model for the prediction of ER when the image information was incorporated. Specifically, for the 2 DL models (ie, the Pre and the PrePost models), the signature of the one with the better AUC on the tuning set was added into the Clinical model to develop the DL_Clinical model using the extra tree algorithm. The details of the model construction were described in the **Supplementary Method 1.2**.

Statistical analysis

In the present study, normally distributed continuous variables were presented with mean ± standard deviation (SD) and compared using Student's *t*-test, and non-normally distributed continuous variables were presented with median (interquartile range) and compared using the Mann-Whitney *U* test. Categorical variables were presented as frequencies with percentages and compared using the chi-squared test. Survival curves were calculated using the Kaplan-Meier method and compared using the log-rank test. According to the Youden index calculated using the tuning cohort, a cutoff for ER prediction score was obtained to stratify the patients into ER or no ER group. The decision curve analysis was used to measure the clinical value of the proposed models.

The area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and F1-score

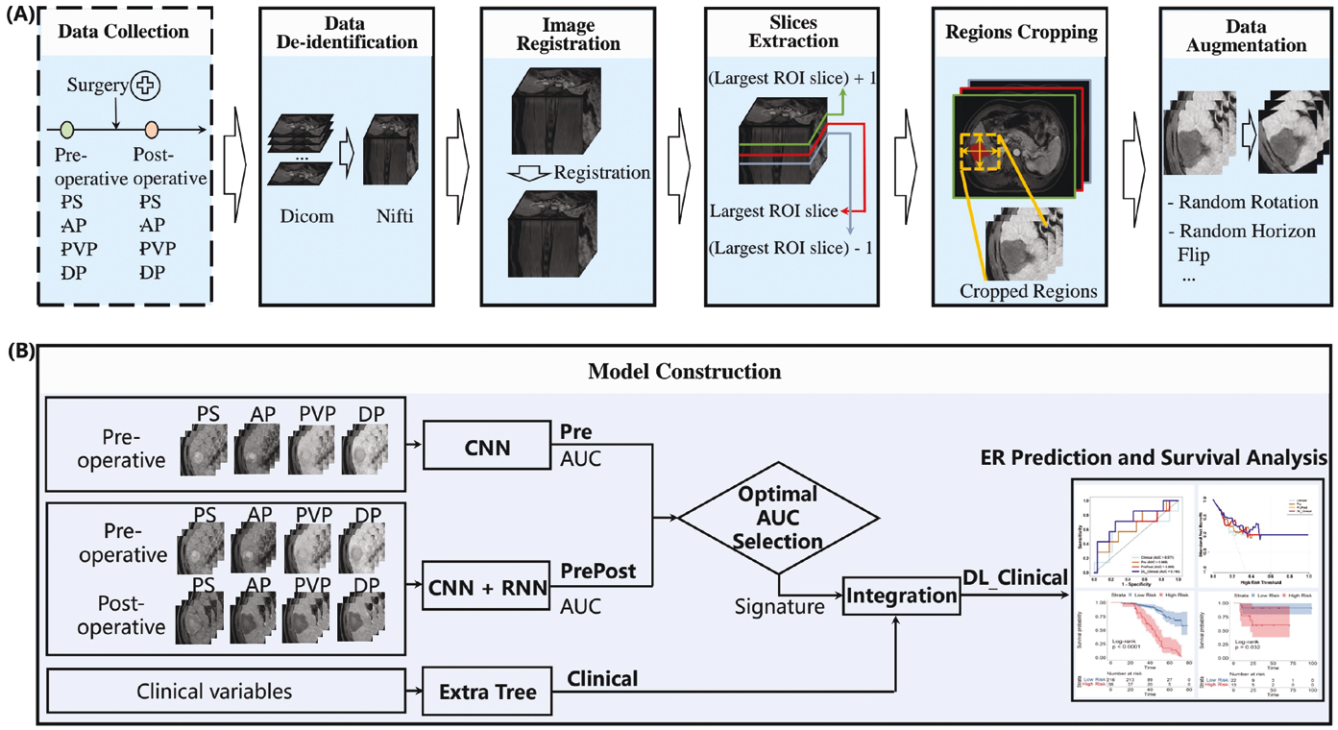

Figure 3. The flowchart of the models for ER prediction. (A) Data collection, PS, AP, PVP, and DP of pre- and post-operative T1WI were collected; Data De-identification, the Dicom format of images was converted into Nifti format; Image Registration, PS, AP, and PVP of T1WI were registered to the DP of T1WI; Slices Extraction, the slice with the largest profile of ROI, and the slices above and below the one with the largest profile of ROI were extracted; Regions Cropping, a refined rectangular region was cropped from each of the extracted slices; Data Augmentation, some data augmentation operations (eg, random rotation, random horizon filp) were performed on the images in the training set; (B) Model Construction, 4 models for ER prediction were constructed. Abbreviations: AP, arterial portal, Clinical, the machine learning model developed using the independent clinical variables; Dicom, digital imaging communications in medicine; DL_Clinical, the machine learning model developed using the independent clinical variables and the signature of PrePost, ER, early recurrence; DP, delayed phase; Nifti, neuroimaging Informatics Technology Initiative; Pre, the deep learning model developed using the pre-operative images; PrePost, the deep learning model developed using the pre- and post-operative images PS, plain scan; PVP, portal vein phase.

with 95% confidence intervals (CIs) were used to evaluate the performance of models in the training cohort and external testing cohorts. The gradient-weight class activation mapping (GRAD-CAM)[22] was used to generate heat maps for the explanations of the DL model.

Statistical analyses were performed using Statistical Package for Social Sciences (SPSS) version 26.0 (IBM Corp., NY, USA), R software version 4.1.2 (http://www.r-project.org/), and Python version 3.8.18 (https://www.python.org/). A 2-tailed *P*-value of less than .05 was considered as statistical significance.

Results

Patient and tumor characteristics

After screening of the patients consecutively treated by TA, a total of 289 patients (41 females and 248 males; mean age, 53.1 ± 10.7 years) were finally recruited in the present study. Specifically, there were 254 patients (33 females and 221 males; mean age, 52.7 ± 10.7 years) from the TAHSYU included in the training cohort. Additionally, in the training cohort, there were 51 (about 20%) patients randomly selected to compose the tuning cohort. Twenty patients (4 females and 16 males; mean age, 54.1 ± 11.2 years) from the SYUCC and 15 patients (4 females and 11 males; mean age, 58.4 ± 10.0 years) from the PHFMU were included in the external testing cohort. The characteristics of the patients in the training cohort and external testing cohort are summarized in Table 1.

Model comparison

The ER prediction performance of the Clinical, Pre, PrePost, and DL_Clinical models were shown in Table 2. Figure 4(A) and (C) shows the ROC curves of the developed models in the training and external testing cohorts, respectively. Figure 4(B) and (D) shows the decision curves analysis of the developed models in the training and external testing cohorts, respectively.

Comparison of the performance between the DL and Clinical models

To explore whether the information in T1WI or the clinical variables was more appropriate for ER prediction, the DL models (ie, the Pre and PrePost models) were compared with the Clinical model. As shown in Table 2, the AUC (0.648, 0.689 vs 0.571), accuracy (0.686, 0.743 vs 0.514), specificity (0.750, 0.786 vs 0.500), PPV (0.300, 0.400 vs 0.222), NPV (0.840, 0.880 vs 0.824), and F1-score (0.353, 0.471 vs 0.320) of DL models (ie, the Pre and PrePost models) were numerically higher than those of the Clinical model in external testing test.

Comparison of the performance between the Pre and PrePost models

To investigate the value of the post-operative T1WI to the ER prediction, the Pre and PrePost models were compared

Table 1. Baseline characteristics of patients with HCC who underwent thermal ablation.

Variables	Training cohorts (n = 254)		*P* value	External testing cohorts (n = 35)		*P* value
	ER (n = 59)	Non-ER (n = 195)		ER (7)	Non-ER (28)	
Demographics						
Mean age (y)	53.6 ± 11.9	52.4 ± 10.2	0.484‡	57.6 ± 10.5	56.5 ± 11.0	0.570‡
Gender			0.769*			0.107*
Female	7 (11.9)	26 (13.3)		0 (0.0)	8 (28.6)	
Male	52 (88.1)	169 (86.7)		7 (100.0)	20 (71.4)	
Comorbidities			0.370*			0.593*
Absence	42 (71.2)	150 (76.9)		4 (57.1)	19 (67.9)	
Presence	17 (28.8)	45 (23.1)		3 (42.9)	9 (32.1)	
HBV			0.605*			0.466*
Absence	3 (5.1)	7 (3.6)		3 (42.9)	8 (28.6)	
Presence	56 (94.9)	188 (96.4)		4 (57.1)	20 (71.4)	
Median ALBI score	-2.8 (-3.0, -2.4)	-2.8 (-3.1, -2.4)	0.595†	-2.8 (-2.9, -2.4)	-2.6 (-2.7, -2.4)	0.430†
Cirrhosis			0.287*			0.593*
Absence	22 (37.3)	88 (45.1)		3 (42.9)	9 (32.1)	
Presence	37 (62.7)	107 (54.9)		4 (57.1)	19 (67.9)	
Tumor feature						
Median tumor size (cm)	2.0 (1.5, 2.6)	2.1 (1.6, 2.6)	0.728†	2.6 (2.0, 3.3)	2.0 (1.5, 2.3)	0.320†
No. of tumor			0.048*			0.042*
Single	41 (69.5)	159 (81.5)		6 (85.7)	28 (100)	
Multiple	18 (30.5)	36 (18.5)		1 (14.3)	0 (0)	
Laboratory findings						
Median AFP (ng/mL)	22.6 (5.6, 102.5)	20.7 (4.4, 165.2)	0.680†	53.5 (28.1, 102.2)	7.2 (2.7, 49.5)	0.146†
Mean ALB (g/L)	40.0 ± 4.1	40.9 ± 4.9	0.469‡	43.6 ± 2.0	42.5 ± 4.3	0.520‡
Median TP (g/L)	69.1 (63.6, 73.2)	67.7 (64.1, 72.8)	0.822†	76.0 (74.5, 86.0)	72.1 (64.0, 81.0)	0.146†
Median AST (U/L)	33.0 (26.5, 40.0)	30.0 (23.0, 40.5)	0.201†	48.0 (42.6, 43.5)	30.7 (24.3, 39.2)	0.004†
Median ALT (U/L)	32.0 (22.0, 45.5)	30.0 (22.0, 44.5)	0.621†	67.0 (53.2, 152.0)	32.4 (29.3, 44.8)	0.001†
Median TBIL (μmol/L)	12.4 (8.7, 16.8)	13.0 (8.7, 18.3)	0.704†	67.7 (14.5, 70.9)	70.0 (12.5, 75.3)	0.531†

Normally distributed continuous variables are presented with mean ± SD, and non-normally distributed continuous variables are presented with median (interquartile range). Data in bracket of category variables are percent of patients.
*P are calculated with χ^2 test.
†P are calculated with Mann-Whitney U test.
‡P are calculated with Student's t-test. Abbreviations: AFP, α-fetoprotein; ALB, albumin; ALBI, albumin-bilirubin; ALT, alanine aminotransferase; AST, aspartate aminotransferase; ER, early recurrence; HBV, viral hepatitis type B; TBIL, total bilirubinl TP, total protein

in the external testing cohort. As shown in **Table 2**, the AUC (0.689 versus 0.648), accuracy (0.743 vs 0.686), sensitivity (0.571 vs 0.429), specificity (0.786 vs 0.750), PPV (0.400 vs 0.300), NPV (0.880 vs 0.840), F1-score (0.471 vs 0.353) of the PrePost model were numerically higher than those of the Pre model.

Comparison between the integrated and non-integrated models

To investigate the improvements of ER prediction performance in the best DL model when the clinical variables were incorporated, an DL_Clinical model was developed. Specifically, because the AUC (0.666 vs 0.551) of the PrePost model was numerically high than that of the Pre model in the tuning set, the signature of the PrePost model was combined with the clinical variables to develop the DL_Clinical model using extra trees algorithm.

In the external testing cohort, as shown in Table 2, for the comparison between the DL_Clinical and Clinical models, the AUC (0.740 vs 0.571), accuracy (0.714 vs 0.514), sensitivity (0.714 vs 0.571), specificity (0.714 vs 0.500), PPV (0.385 vs 0.222), NPV (0.909 vs 0.824), and F1-score (0.500 vs 0.320) of the DL_Clinical model were higher than those of the Clinical model. For the comparison between the DL_Clinical and PrePost models, although the accuracy (0.743 vs 0.714), specificity (0.786 vs 0.714), and PPV (0.400 vs 0.385) of the PrePost model were slightly higher than those of DL_Clinical model, the total performance of the DL_Clinical model was numerically superior to the PrePost model when comprehensively considering the AUC (0.740 vs 0.689), sensitivity (0.714 vs 0.571), NPV (0.909 vs 0.880), and F1-score (0.500 vs 0.471), particularly the great improvement on sensitivity. Additionally, as shown in Figure 4(D), in the external testing cohort, the decision curve analysis graphically demonstrated that the DL_Clinical model provided a larger net benefit across the range of reasonable threshold probabilities compared with the Clinical and PrePost models. In total, the DL_Clinical model outperformed the Clinical model and the PrePost model in the external testing cohort.

Table 2. The performance of the models in the training and external testing cohorts.

Cohort	Models	AUC	Accuracy	Sensitivity	Specificity	PPV	NPV	F1-score
Training cohort	Clinical	0.778 [0.710, 0.843]	0.807 [0.756, 0.858]	0.169 [0.083, 0.267]	1.000 [1.000, 1.000]	1.000 [1.000, 1.000]	0.799 [0.748, 0.852]	0.290 [0.154, 0.421]
	Pre	0.872 [0.734, 0.849]	0.764 [0.709, 0.815]	0.847 [0.746, 0.935]	0.738 [0.679, 0.800]	0.495 [0.396, 0.593]	0.941 [0.901, 0.975]	0.625 [0.527, 0.712]
	PrePost	0.854 [0.667, 0.758]	0.587 [0.528, 0.646]	0.949 [0.887, 1.000]	0.477 [0.412, 0.550]	0.354 [0.277, 0.430]	0.969 [0.929, 1.000]	0.516 [0.428, 0.596]
	DL_Clinical	0.931 [0.900, 0.958]	0.878 [0.839, 0.917]	0.559 [0.441, 0.686]	0.974 [0.949, 0.995]	0.868 [0.744, 0.969]	0.880 [0.834, 0.925]	0.680 [0.568, 0.777]
External testing cohort	Clinical	0.571 [0.253, 0.885]	0.514 [0.343, 0.686]	0.571 [0.200, 1.000]	0.500 [0.310, 0.704]	0.222 [0.059, 0.444]	0.824 [0.625, 1.0]	0.320 [0.091, 0.545]
	Pre	0.648 [0.384, 0.907]	0.686 [0.543, 0.857]	0.429 [0.000, 0.833]	0.750 [0.613, 0.923]	0.300 [0.000, 0.667]	0.840 [0.692, 0.964]	0.353 [0.000, 0.636]
	PrePost	0.689 [0.368, 0.935]	0.743 [0.600, 0.886]	0.571 [0.167, 1.000]	0.786 [0.621, 0.926]	0.400 [0.111, 0.714]	0.880 [0.731, 1.000]	0.471 [0.143, 0.714]
	DL_Clinical	0.740 [0.462, 0.935]	0.714 [0.543, 0.857]	0.714 [0.333, 1.000]	0.714 [0.519, 0.871]	0.385 [0.133, 0.667]	0.909 [0.760, 1.000]	0.500 [0.200, 0.727]

Abbreviations: AUC, area under the receiver operating characteristic curve; Clinical, the machine learning model developed using the independent clinical variables; DL_Clinical, the machine learning model developed using the independent clinical variables and the signature of PrePost; NPV, negative predictive value; PPV, positive predictive value; Pre, the deep learning model developed using the pre-operative images; PrePost, the deep learning model developed using the pre- and post-operative images.

Visualization for the DL model

The GRAD-CAM was applied to the PrePost model to generate the heatmaps for the original images. The heatmaps and the original maps of 2 patient examples are shown in Supplementary Figure S1. These 2 examples are both patients with ER. However, the left example was predicted as patient with ER, whereas the right example was predicted as no patient with ER. There were obvious differences in the heatmaps between the 2 examples. For the left example, the high intensity regions (red areas) in the post-operative heatmaps mainly were the edges of ablation zones, not the tumors, possibly because the tumor cells were destroyed using the thermal ablation. Curiously, the high intensity regions in the pre-operative maps mainly were the edges of tumors as well. This may be attributed to the fact that the pre- and post-operative images were simultaneously used to train the model. The attention of the model may be affected due to the edges of ablation zones. For the right example, although the high intensity regions in the pre-operative heatmaps included the tumors, the edges of ablation zones in the post-operative images were not be focused.

The features importance of the integrated model

To investigate the importance of each feature used to develop the DL_Clinical model, the mean decrease in impurity[23] was used to evaluate the feature importance. As shown in Supplementary Table S2, the PrePost signature accounted for a large percentage (0.764) in total features. This indicated that the information in T1WI (ie, the PrePost signature) was more important than the clinical variables for ER prediction.

Association between DL-predicted ER and survival

The patients with HCC were stratified into low-risk group and high-risk group based on the cutoff for ER prediction score. As shown in **Figure 5** (A)–(C), for the Clinical, PrePost, and DL_Clinical models, the low-risk group and high-risk group showed significantly different RFS ($P < .01$, $P < .01$, and $P < .01$) in the training cohort, respectively. However, as shown in Figure 5(D) and (E), in the external testing cohort, there were no significant difference in RFS between the low- and high-risk groups which were divided by the Clinical and PrePost models ($P = .71$ and $P = .06$), respectively. As shown in Figure 5 (F), only the DL_Clinical model can stratify patients into low-risk group and high-risk group with a significant difference of RFS ($P = .04$) in the external testing cohort.

Discussion

In the present study, several models were developed using T1WI to predict ER for patients with HCC who underwent TA, and the integrated model (ie, the DL_Clinical model) which incorporated the PrePost signature and clinical variables yielded the optimal ER predictive performance in the external testing cohort. The comparisons among the models were conducted. Additionally, the model interpretation and risk stratification were implemented in the present study.

To date, several MRI studies have been conducted to predict the ER for patients with HCC who underwent TA using artificial intelligence (AI) algorithm. For example, Huang et al.[24] and Iseke et al.[25] developed radiomics models based on pre-operative MRI to predict the ER for patients with HCC after TA, and their radiomics models achieved the AUCs of 0.77 and 0.75 in testing sets, respectively. However, only pre-operative MRI was used in their studies,[24,25] and thus the prediction performance of their models may be limited due to lack of the abundant physiological information in post-operative MRI. Based on the experimental results in the present study, the post-operative MRI can improve the ER predictive performance of the Pre model which was developed only using pre-operative MRI. Additionally, in these 2 studies,[24,25] the patients were collected from a single center, and

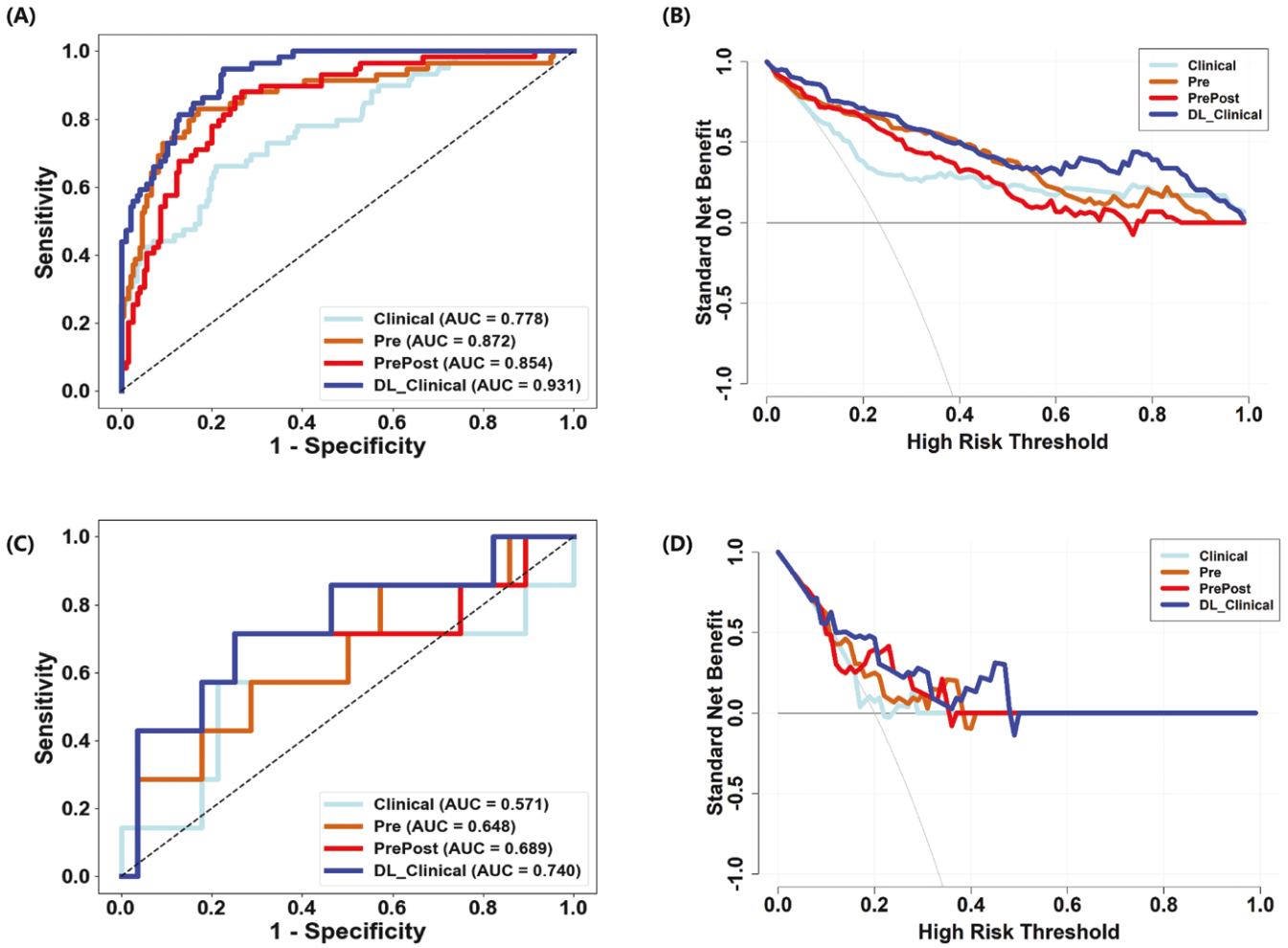

Figure 4. The ROC curves and decision curve analysis of the models in the training cohort and external testing cohort, respectively. (A) The ROC curves of the models in the training cohort. (B) The decision curve analysis of the models in the training cohort. (C) The ROC curves of the models in the external testing cohort. (D) The decision curve analysis of the models in the external testing cohort. Abbreviations: Clinical, the machine learning model developed using the independent clinical variables; Pre, the deep learning model developed using the pre-operative images; PrePost, the deep learning model developed using the pre- and post-operative images; DL_Clinical, the machine learning model developed using the independent clinical variables and the signature of PrePost.

the generalization of their models may need to be validated in independent external testing sets. In contrast, all models in our experiments were validated in an external testing set, and the generalization of the models can be assessed.

In the present study, the longitudinal sequences (images at pre- and post-operative) were used to develop the PrePost model. Therefore, in addition to the static semantic information in the images, there was a dynamic time-dependent relation between the images at the 2 time points (ie, pre- and post-operative). Therefore, the PrePost model was developed using 2 types of neural networks (ie, CNN[26] and RNN[27]). The CNN was used to extract features of T1WI at each time point, and the RNN was used to learn the relation between pre- and post-operative features for the ER prediction.

The present study found that the performance of each DL model (the Pre or PrePost models) was higher than that of the Clinical model in the external testing cohort. This suggested that the subtle information which cannot be captured by naked eye in images may be more important than clinical variables for the ER prediction. This finding was consistent with those of Shen et al[28] and Yuan et al,[29] who found that the ER prediction performance of the models developed using clinical variables was lower than that of the models developed using images. The present study found that each of the evaluation results of the PrePost model was higher than that of the Pre model in the external testing cohort. This suggested that post-operative T1WI can provide the information which was not in pre-operative T1WI, for example the information of the edges of ablation zones, and this information was important to predict ER for patients with HCC. This finding was consistent with that of Beleu et al,[30] who found that the information of ablation zones in post-operative images was critical to predict the recurrence of patients with HCC. Additionally, the present study found that the DL_Clinical model which combined the image information (ie, PrePost signature) and the clinical variables achieved the best performance, indicating that the 2 types of information were complementary. Similarly, Kim et al[31] found that the model developed using the radiomics features and clinical variables can provide a better survival estimation of patients with HCC than the clinical-only model and radiomics-only model.

The heatmap visualization and feature importance analysis were conducted for the interpretations of the PrePost model and DL_Clinical model, respectively. For the heatmap

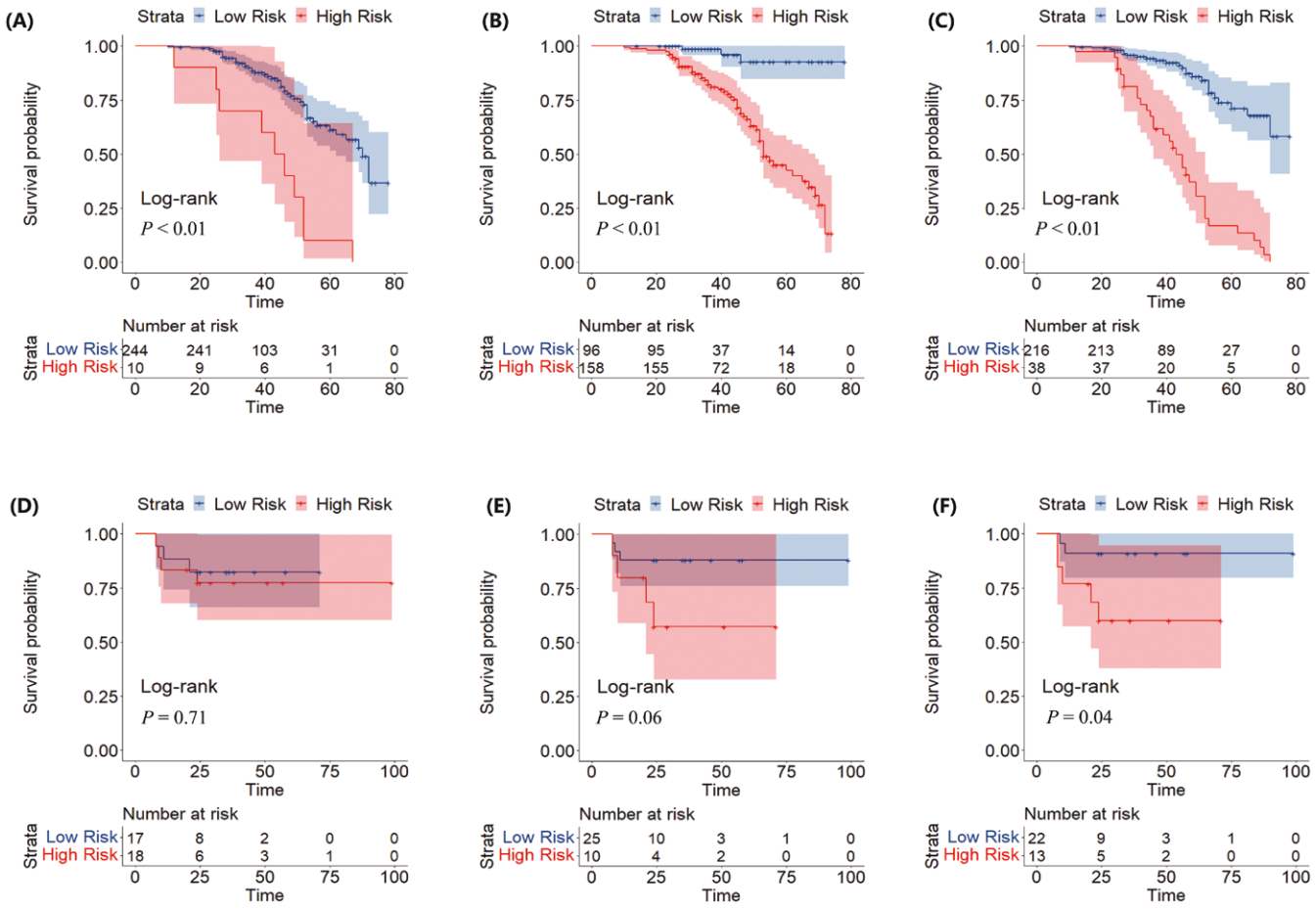

Figure 5. The RFS stratification of the Clinical, PrePost, and DL_Clinical models in the training cohort and external testing cohort. (A) RFS of the patients who were stratified based on the Clinical model in the training cohort. (B) RFS of the patients who were stratified based on the PrePost model in the training cohort. (C) RFS of the patients who were stratified based on the DL_Clinical model in the training cohort. (D) RFS of the patients who were stratified based on the Clinical model in the external testing cohort. (E) RFS of the patients who were stratified based on the PrePost model in the external testing cohort. (F) RFS of the patients who were stratified based on the DL_Clinical model in the external testing cohort. Abbreviations: Clinical, the machine learning model developed using the independent clinical variables; DL_Clinical, the machine learning model developed using the independent clinical variables and the signature of PrePost; ER, early recurrence; PrePost, the deep learning model developed using the pre- and post-operative images; RFS, recurrence free survival.

visualization for the PrePost model, the tumors and edges of ablation zones were important for the correct ER prediction. Additionally, for the feature importance analysis for the DL_Clinical model, the present study found that the information in T1WI (ie, PrePost signature) was more important than the clinical variables for ER prediction. Therefore, through the interpretations of the models, radiologists and physicians may get profound understandings to the ER mechanism of patients with HCC.

The PrePost model stratify patients into low-risk group and high-risk group with a marginal significant difference in RFS (P = .06) in the external testing cohort. After the clinical variables were integrated with the signature of the PrePost model, the DL_Clinical model can stratify HCC patients into low-risk group and high-risk group with a significant difference in RFS (P = .04) in the external testing cohort. This suggested that the clinical variables can improve the performance of the PrePost model, and the information in T1WI (i.e., PrePost signature) and clinical variables were complementary for the ER prediction of patients with HCC. Additionally, the DL_Clinical model may be an appropriate tool that can help doctors to select personalized surveillance strategy for patients with HCC. Specifically, low-risk patients may receive no preventive TACE and a less intensive surveillance regimen, even within the first 2 years after TA. This less intensive surveillance regimen can reduce healthcare cost,[32] alleviate patients' anxiety,[33] and minimize the risk of overtreatment.[34] High-risk patients should undergo intensive surveillance lasting as much as possible because of the high risk for recurrence even after 2 years, accompanied by adjuvant systemic therapies.

There are certain limitations in our study. First, the tumors and ablation zones were delineated manually by expert radiologists. This was a time-consuming procedure. Additionally, the inaccurate delineation may affect the performance of models. Therefore, an automated model which can segment tumors accurately needs to be developed. Second, the information of medical records or gene was not available in this study. The performance of the models may be improved greatly when these variables were added. Third, the analysis of distinguishing between the lesions that were away from and those that were abutting ablated areas may not be conducted due to the relatively small number of patients with ER in the present study. In future, more patients need to be collected to develop a DL model for the distinguishing between the 2 types of the

lesions, guiding the interventional radiologists to make further treatment plans.

In conclusion, the present study developed the ER prediction models based on longitudinal MRI. The integration model (ie, DL_Clinical) showed a high-level of ER predictive performance in the external testing cohort, which may help doctors in surveillance strategy selection for patients with HCC who underwent TA.

Acknowledgments

The authors would like to thank the medical staff who supported the research.

Author contributions

Kai Li, Qingyang Kong (Conceptualization). Kai Li (Data curation). Qingyang Kong (Data curation). Kai Li (Formal Analysis). Kai Li, Qingyang Kong (Writing—original draft). Kai Li, Qingyang Kong (Writing—review & editing).

Funding

None declared.

Conflict of Interest

The authors indicated no financial relationships.

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Supplementary material

Supplementary material is available at *The Oncologist* online.

References

1. Forner A, Reig M, Bruix J. Hepatocellular carcinoma. *Lancet*. 2018;391:1301-1314. https://doi.org/10.1016/S0140-6736(18)30010-2
2. Villanueva A. Hepatocellular carcinoma. *N Engl J Med*. 2019;380:1450-1462. https://doi.org/10.1056/NEJMra1713263
3. Yang JD, Hainaut P, Gores GJ, et al. A global view of hepatocellular carcinoma: trends, risk, prevention and management. *Nat Rev Gastroenterol Hepatol*. 2019;16:589-604. https://doi.org/10.1038/s41575-019-0186-y
4. Heimbach JK, Kulik LM, Finn RS, et al. AASLD guidelines for the treatment of hepatocellular carcinoma. *Hepatology*. 2018;67:358-380. https://doi.org/10.1002/hep.29086
5. Liver EAS. EASL clinical practice guidelines: management of hepatocellular carcinoma. *J Hepatol*. 2018;69:182-236.
6. Kim GA, Shim JH, Kim MJ, et al. Radiofrequency ablation as an alternative to hepatic resection for single small hepatocellular carcinomas. *Br J Surg*. 2016;103:126-135. https://doi.org/10.1002/bjs.9960
7. Ward EM, Sherif AE, O'Neill S, et al. Clinical outcomes of ablation compared with resection for single hepatocellular carcinoma lesions, as a primary treatment or bridging to liver transplantation: a retrospective comparative study. *Ann Transplant*. 2021;26:e931980. https://doi.org/10.12659/AOT.931980
8. Izzo F, Granata V, Grassi R, et al. Radiofrequency ablation and microwave ablation in liver tumors: an update. *Oncologist*. 2019;24:e990-e1005. https://doi.org/10.1634/theoncologist.2018-0337
9. Altekruse SF, McGlynn KA, Dickie LA, Kleiner DE. Hepatocellular carcinoma confirmation, treatment, and survival in surveillance, epidemiology, and end results registries, 1992-2008. *Hepatology*. 2012;55:476-482. https://doi.org/10.1002/hep.24710
10. Portolani N, Coniglio A, Ghidoni S, et al. Early and late recurrence after liver resection for hepatocellular carcinoma—prognostic and therapeutic implications. *Ann Surg*. 2006;243:229-235. https://doi.org/10.1097/01.sla.0000197706.21803.a1
11. Kim S, Shin J, Kim DY, et al. Radiomics on gadoxetic acid-enhanced magnetic resonance imaging for prediction of postoperative early and late recurrence of single hepatocellular carcinoma. *Clin Cancer Res*. 2019;25:3847-3855. https://doi.org/10.1158/1078-0432.CCR-18-2861
12. Villanueva A, Hoshida Y, Battiston C, et al. Combining clinical, pathology, and gene expression data to predict recurrence of hepatocellular carcinoma. *Gastroenterology*. 2011;140:1501-1512. https://doi.org/10.1053/j.gastro.2011.02.006
13. Notarpaolo A, Layese R, Magistri P, et al. Validation of the AFP model as a predictor of HCC recurrence in patients with viral hepatitis-related cirrhosis who had received a liver transplant for HCC. *J Hepatol*. 2017;66:552-559. https://doi.org/10.1016/j.jhep.2016.10.038
14. Liu F, Liu D, Wang K, et al. Deep learning radiomics based on contrast-enhanced ultrasound might optimize curative treatments for very-early or early-stage hepatocellular carcinoma patients. *Liv Cancer*. 2020;9:397-413. https://doi.org/10.1159/000505694
15. Kang TW, Lim HK, Lee MW, et al. Aggressive intrasegmental recurrence of hepatocellular carcinoma after radiofrequency ablation: risk factors and clinical significance. *Radiology*. 2015;276:274-285. https://doi.org/10.1148/radiol.15141215
16. Aerts HJWL, Velazquez ER, Leijenaar RTH, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*. 2014;5:4006. https://doi.org/10.1038/ncomms5006
17. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25:44-56. https://doi.org/10.1038/s41591-018-0300-7
18. Galle PR, Forner A, Llovet JM, et al. EASL clinical practice guidelines: management of hepatocellular carcinoma. *J Hepatol*. 2018;69:182-236. https://doi.org/10.1016/j.jhep.2018.03.019
19. Imamura H, Matsuyama Y, Tanaka E, et al. Risk factors contributing to early and late phase intrahepatic recurrence of hepatocellular carcinoma after hepatectomy. *J Hepatol*. 2003;38:200-207. https://doi.org/10.1016/s0168-8278(02)00360-4
20. Yang XZ, Yuan CW, Zhang YH, Li K, Wang Z. Predicting hepatocellular carcinoma early recurrence after ablation based on magnetic resonance imaging radiomics nomogram. *Medicine (Baltimore)*. 2022;101:e32584. https://doi.org/10.1097/md.0000000000032584
21. He LL, Liu XL, Zhang S, et al. Independent risk factors for disease recurrence after surgery in patients with hepatitis B virus-related hepatocellular carcinoma≤ 3 cm in diameter. *Gastroenterol Rep*. 2019;7:250-257. https://doi.org/10.1093/gastro/goz009
22. Selvaraju RR, Cogswell M, Das A, et al. Grad-cam: visual explanations from deep networks via gradient-based localization. *Proceed IEEE Int Conf Comput Vision*. 2017;618-626. https://doi.org/10.1109/iccv.2017.74
23. Breiman L. *Manual on Setting Up, Using, and Understanding Random Forests v3. 1*. Statistics Department University of California Berkeley; 2002.
24. Huang WR, Pan YF, Wang HF, et al. Delta-radiomics analysis based on multi-phase contrast-enhanced MRI to predict early recurrence in hepatocellular carcinoma after percutaneous thermal ablation. *Acad Radiol*. 2024;31:4934-4945. https://doi.org/10.1016/j.acra.2024.06.002
25. Iseke S, Zeevi T, Kucukkaya A, et al. Machine learning models for prediction of posttreatment recurrence in early-stage hepatocellular carcinoma using pretreatment clinical and MRI features: a proof-of-concept study. *Am J Roentgenol*. 2023;220:245-255.

26. Zhao LT, Bao J, Qiao XM, et al. Predicting clinically significant prostate cancer with a deep learning approach: a multicentre retrospective study. *Eur J Nucl Med Mol Imaging*. 2023;50:727-741.
27. Lu L, Dercle L, Zhao BS, Schwartz LH. Deep learning for the prediction of early on-treatment response in metastatic colorectal cancer from serial medical imaging. *Nat Commun*. 2021;12:6654. https://doi.org/10.1038/s41467-021-26990-6
28. Shen JX, Zhou Q, Chen ZH, et al. Longitudinal radiomics algorithm of posttreatment computed tomography images for early detecting recurrence of hepatocellular carcinoma after resection or ablation. *Transl Oncol*. 2021;14:100866. https://doi.org/10.1016/j.tranon.2020.100866
29. Yuan CW, Wang ZC, Gu DS, et al. Prediction early recurrence of hepatocellular carcinoma eligible for curative ablation using a Radiomics nomogram. *Cancer Imaging*. 2019;19:1-12.
30. Beleù A, Autelitano D, Geraci L, et al. Radiofrequency ablation of hepatocellular carcinoma: CT texture analysis of the ablated area to predict local recurrence. *Eur J Radiol*. 2022;150:110250. https://doi.org/10.1016/j.ejrad.2022.110250
31. Kim J, Choi SJ, Lee SH, Lee HY, Park H. Predicting survival using pretreatment CT for patients with hepatocellular carcinoma treated with transarterial chemoembolization: comparison of models using radiomics. *Am J Roentgenol*. 2018;211:1026-1034. https://doi.org/10.2214/ajr.18.19507
32. Thompson Coon J, Rogers G, Hewson P, et al. Surveillance of cirrhosis for hepatocellular carcinoma: a cost–utility analysis. *Br J Cancer*. 2008;98:1166-1175. https://doi.org/10.1038/sj.bjc.6604301
33. Narasimman M, Hernaez R, Cerda V, et al. Hepatocellular carcinoma surveillance may be associated with potential psychological harms in patients with cirrhosis. *Hepatology*. 2024;79:107-117. https://doi.org/10.1097/HEP.0000000000000528
34. Atiq O, Tiro J, Yopp AC, et al. An assessment of benefits and harms of hepatocellular carcinoma surveillance in patients with cirrhosis. *Hepatology*. 2017;65:1196-1205. https://doi.org/10.1002/hep.28895