ORIGINAL ARTICLE

OPEN

An Integrated Model Combined Conventional Radiomics and Deep Learning Features to Predict Early Recurrence of Hepatocellular Carcinoma Eligible for Curative Ablation: A Multicenter Cohort Study

Yong-hai Li, PhD,*†‡ Gui-xiang Qian, MD,† Yu Zhu, MD,§ Xue-di Lei, MD,‖
Lei Tang, MD,¶# Xiang-yi Bu, MD,† Ming-tong Wei, MD,¶# and
Wei-dong Jia, PhD*†

Objective: Hepatocellular carcinoma (HCC) is the most common primary liver malignancy. Ablation therapy is one of the first-line treatments for early HCC. Accurately predicting early recurrence (ER) is crucial for making precise treatment plans and improving prognosis. This study aimed to develop and validate a model (DLRR) that incorporates deep learning radiomics and traditional radiomics features to predict ER following curative ablation for HCC.

Methods: We retrospectively analysed the data of 288 eligible patients from 3 hospitals—1 primary cohort (center 1, n = 222) and 2 external test cohorts (center 2, n = 32 and center 3, n = 34)—from April 2008 to March 2022. 3D ResNet-18 and PyRadiomics were applied to extract features from contrast-enhanced computed tomography (CECT) images. The 3-step (ICC-LASSO-RFE) method was used for feature selection, and 6 machine learning methods were used to construct models. Performance was compared through the area under the receiver operating characteristic curve (AUC), net reclassification improvement (NRI) and integrated discrimination improvement (IDI) indices. Calibration and clinical applicability were assessed through calibration curves and decision curve analysis (DCA), respectively. Kaplan-Meier (K-M) curves were generated to stratify patients based on progression-free survival (PFS) and overall survival (OS).

Results: The DLRR model had the best performance, with AUCs of 0.981, 0.910, and 0.851 in the training, internal validation, and external validation sets, respectively. In addition, the calibration curve and DCA curve revealed that the DLRR model had good calibration ability and clinical applicability. The K-M curve indicated that the DLRR model provided risk stratification for progression-free survival (PFS) and overall survival (OS) in HCC patients.

Conclusions: The DLRR model noninvasively and efficiently predicts ER after curative ablation in HCC patients, which helps to categorize the risk in patients to formulate precise diagnosis and treatment plans and management strategies for patients and to improve the prognosis.

Key Words: HCC, ablation, early recurrence, radiomics, deep learning

(J Comput Assist Tomogr 2025;49:860–871)

Primary liver cancer is the second leading cause of cancer-related death globally, and its incidence is increasing. The number of new patients with primary liver cancer worldwide is expected to exceed one million by 2025.[1,2] Hepatocellular carcinoma (HCC) accounts for ~90% of primary liver cancers and poses a significant challenge to global health because of its poor treatment outcomes.[3] Ablation has been shown to be a minimally invasive and first-line treatment for early-stage HCC, achieving therapeutic outcomes similar to those of surgical resection, but the recurrence rate within 5 years is ~70%, which is higher than that of surgical resection, limiting its efficacy.[4–7] Early recurrence (ER), typically defined as recurrence within 2 years after curative treatment for HCC, leads to increased mortality and a poorer survival prognosis compared with late recurrence.[8–10] Since HCC patients are mostly diagnosed by imaging without histopathologic examination and lack risk factors related to recurrence, such as microvascular invasion (MVI),[11] predicting ER after ablation has been a difficult clinical research task. Therefore, there is an urgent need for a reliable technique that can noninvasively and efficiently predict ER after ablation in HCC patients to

formulate a precise diagnostic plan for patients as well as a management strategy to improve prognosis.[11,12]

Radiomics was initially proposed in 2012 as a comprehensive method for analyzing medical images to quantify imaging phenotypes and advance precision medicine.[13] In recent years, predictive models based on radiomics have become increasingly popular for assessing the risk of HCC recurrence after surgical resection.[14–17] Qian constructed an integrated model combining radiomic features with clinicoradiologic features, which demonstrated effective predictive performance for ER after liver resection for HCC in patients with cirrhosis.[18] However, predictive models for ER after ablation therapy for HCC remain limited in both quantity and clinical translation. Yuan et al developed a radiomics nomogram integrating radiomic features from multiphase contrast-enhanced computed tomography (CECT) and clinicopathologic data, achieving a C-index of 0.755 in validation.[19] However, their approach was limited by extracting features solely from the tumor region, without considering the potentially informative peritumoral tissue. Another study by Peng et al[20] constructed a radiomics-based model using both tumor and peritumoral features, demonstrating good predictive performance (*C*-index: 0.801). Nonetheless, their approach did not incorporate deep-learning methods, which may enhance predictive accuracy and robustness.

Deep-learning has great application prospects in the field of medical imaging. Li et al[21] introduced deep-learning-based radiomics (DLR), demonstrating its superiority over traditional radiomics in predictive performance. Convolutional neural networks (CNNs), particularly ResNet-18, with its lightweight architecture and efficient feature extraction capability, making it suitable for clinical medical imaging analysis. Wang et al developed a ResNet-18-based multimodal model using preoperative magnetic resonance imaging (MRI) and CT images to predict MVI in HCC, achieving an AUC of 0.819.[22] Nevertheless, deep-learning approaches face challenges such as limited data sets and overfitting. The combination of conventional radiomics and deep-learning features can improve model accuracy and robustness. Wei et al integrated deep-learning and traditional radiomics (DLRR) to predict muscle invasion in bladder cancer, attaining high predictive accuracy.[23] However, the application of DLRR to predict ER after curative ablation in HCC remains limited.

In this study, we aimed to construct a model based on CECT images through the DLRR method to predict ER in HCC patients following curative ablation to assist in the development of early-stage HCC treatment and prognostic management programs and provide a feasible solution for the application of deep-learning to the ablation treatment of early-stage HCC.

MATERIALS AND METHODS

Patient Selection

This multicenter retrospective cohort study was approved by the Ethics Management Committee of the First Affiliated Hospital of the University of Science and Technology of China(2021-RE-043). The institutional ethics review board has approved our study, and the requirement for informed consent was waived because of the retrospective nature of the study. We enrolled a total of 288 patients with early-stage HCC who received curative ablation at 3 centers: the First Affiliated Hospital of the University of Science and Technology of China (Center 1), the Second Hospital of Anhui Medical University (Center 2), and the Taizhou Hospital of Zhejiang Province (Center 3) from April 2008 to March 2022. The inclusion and exclusion criteria for patient selection are shown in Fig. 1. At Center 1, patients were randomly assigned to a training set and an internal validation set at a 7:3 ratio. Patients from Center 2 and Center 3 were combined to form the external validation set. In this study, the design, implementation, and
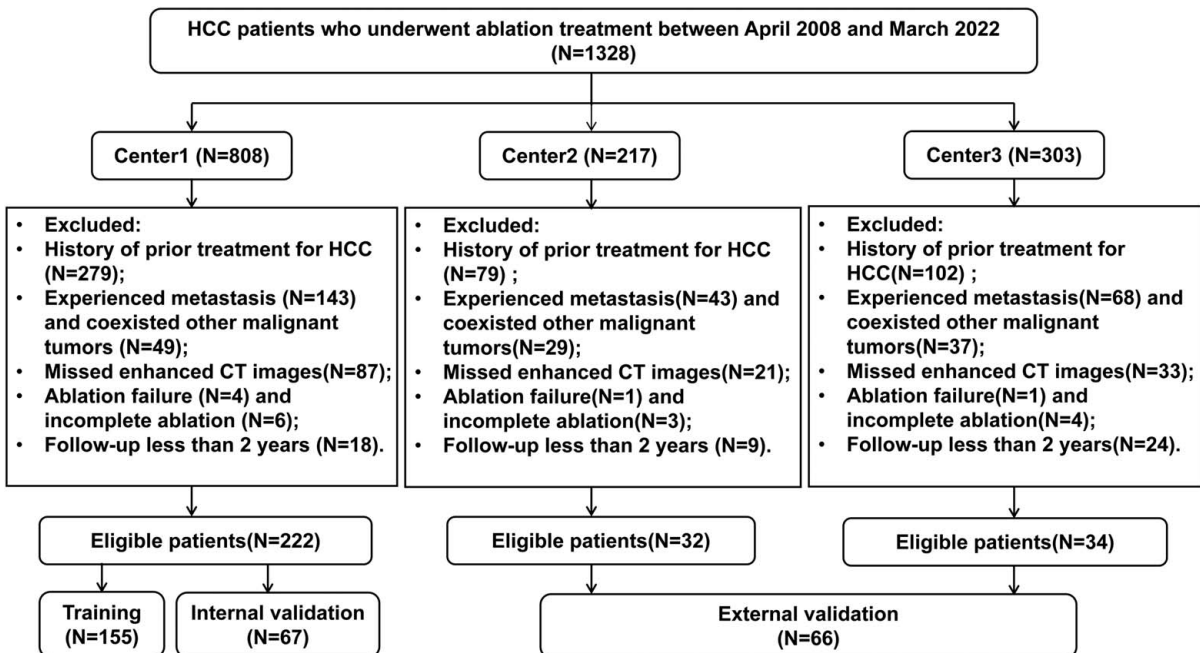

FIGURE 1. Patient inclusion and exclusion flowchart.

reporting of radiomics analysis were conducted according to the Checklist for Evaluating Radiomics Applications (CLEAR) guidelines.[24]

Ablation was performed by the same experienced team of the respective hospitals under ultrasound guidance as described in eAppendix 1, Supplemental Digital Content 1, http://links.lww.com/JCAT/A26. If there were multiple lesions, ablation was performed one by one. All patients underwent CECT, and the imaging data were collected through the picture archiving and communication system (PACS) of the respective hospitals. Specific CT equipment information is shown in eAppendix 2, Supplemental Digital Content 1, http://links.lww.com/JCAT/A26. The inclusion criteria were as follows: (1) clinical diagnosis of HCC according to the noninvasive criteria established by the American Association for the Study of Liver Disease based on distinct imaging features;[25] (2) single tumor diameter ≤ 5 cm, multiple tumors ≤ 3, each diameter ≤ 3 cm; (3) refusal to undergo hepatectomy or liver transplantation; and (4) patient management involving curative ablation only. The exclusion criteria were as follows: (1) tumors invading blood vessels, bile ducts, adjacent organs, distant metastasis, or other malignancies; (2) a prior history of HCC treatment, such as hepatic resection, transarterial chemoembolization (TACE), targeted therapy, or radiotherapy; (3) the absence of CECT imaging data or CECT conducted more than 1 month before ablation; and (4) a follow-up duration of < 2 years. The detailed workflow is illustrated in Figure 2.

Clinical Data Collection

Before initiating the data collection process, all relevant staff members at the participating centers underwent a training session on data extraction. The medical records of eligible patients were meticulously reviewed. Clinical and laboratory data were systematically collected using standardized forms with predefined fields for demographics, tumor characteristics, laboratory results (such as liver function and blood counts), treatment details, and other clinical parameters to ensure consistent data collection across centers. To ensure relevance and consistency, only the data obtained within 1 week before ablation were considered. To enhance data accuracy and minimize extraction errors, another reviewer independently assessed and validated 30% of the collected data. To account for potential variations across participating centers, all laboratory measurements were standardized. Extreme outliers—values significantly higher or lower than the norm—were flagged for review. These outliers were re-evaluated by the hospital's lead researcher or the designated chief physician to confirm their validity and exclude input errors.

Image Information Collection

CECT images were obtained from the hospital PACS in DICOM format. Two physicians (reader 1 and reader 2) from each center independently evaluated the CECT images and focused on the following 8 semantic features: (1) tumor margin; (2) tumor capsule; (3) intratumoral vessels; (4) tumor growth; (5) intratumoral necrosis; and (6)
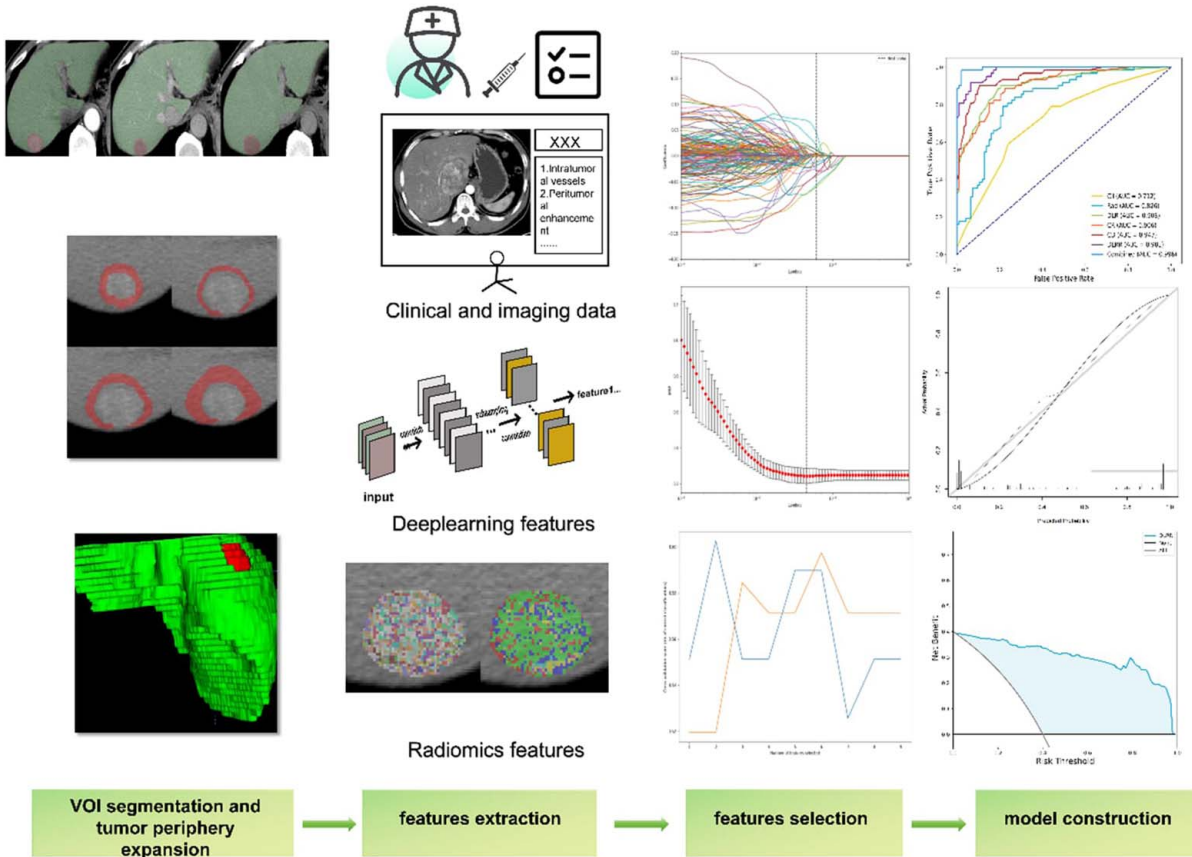

FIGURE 2. The specific research process.

TABLE 1. Clinical and Image Characteristics of 288 Patients in Different Cohorts

	Training (N = 155), n (%)	Internal validation (N = 67), n (%)	External validation (N = 66), n (%)	*P*
Tumor size (cm), mean ± SD	2.72 ± 1.05	2.70 ± 1.10	2.57 ± 0.90	0.586
Age				0.625
43	13 (8.4)	8 (11.9)	5 (7.58)	
≥ 43	142 (91.6)	59 (88.1)	61 (92.4)	
History of hepatitis				0.055
Absent	61 (39.4)	25 (37.3)	15 (22.7)	
Present	94 (60.6)	42 (62.7)	51 (77.3)	
Cirrhosis				0.424
Absent	28 (18.1)	17 (25.4)	12 (18.2)	
Present	127 (81.9)	50 (74.6)	54 (81.8)	
NLR				0.769
< 2.66	123 (79.4)	54 (80.6)	50 (75.8)	
≥ 2.66	32 (20.6)	13 (19.4)	16 (24.2)	
PLR				0.202
< 65.217	68 (43.9)	30 (44.8)	21 (31.8)	
≥ 65.217	87 (56.1)	37 (55.2)	45 (68.2)	
ALT (U/L)				0.314
< 98	152 (98.1)	63 (94.0)	64 (97.0)	
≥ 98	3 (1.9)	4 (6.0)	2 (3.0)	
AST (U/L)				0.486
< 22	15 (9.7)	7 (10.4)	10 (15.2)	
≥ 22	140 (90.3)	60 (89.6)	56 (84.8)	
GGT (U/L)				0.774
< 219	143 (92.3)	60 (89.6)	61 (92.4)	
≥ 219	12 (7.7)	7 (10.4)	5 (7.6)	
TB (μmol/L)				0.663
< 10.5	17 (11.0)	10 (14.9)	7 (10.6)	
≥ 10.5	138 (89.0)	57 (85.1)	59 (89.4)	
FIB (g/L)				0.087
< 2.22	90 (58.1)	31 (46.3)	29 (43.9)	
≥ 2.22	65 (41.9)	36 (53.7)	37 (56.1)	
INR				0.615
< 1.2	101 (65.2)	48 (71.6)	43 (65.2)	
≥ 1.2	54 (34.8)	19 (28.4)	23 (34.8)	
HbsAg				0.319
Absent	29 (18.7)	12 (17.9)	7 (10.6)	
Present	126 (81.3)	55 (82.1)	59 (89.4)	
AFP (ng/ml)				0.242
< 36.67	79 (51.0)	33 (49.3)	41 (62.1)	
≥ 36.67	76 (49.0)	34 (50.7)	25 (37.9)	
Sex				0.386
Female	30 (19.4)	13 (19.4)	18 (27.3)	
Male	125 (80.6)	54 (80.6)	48 (72.7)	
Capsule appearance				0.072
Absent	100 (64.5)	36 (53.7)	48 (72.7)	
Present	55 (35.5)	31 (46.3)	18 (27.3)	
Intratumor vascularity				0.438
Absent	63 (40.6)	29 (43.3)	33 (50.0)	
Present	92 (59.4)	38 (56.7)	33 (50.0)	
Tumor growth patter				0.055
Absent	120 (77.4)	46 (68.7)	41 (62.1)	
Present	35 (22.6)	21 (31.3)	25 (37.9)	
Intratumor necrosis				0.187
Absent	86 (55.5)	45 (67.2)	35 (53.0)	
Present	69 (44.5)	22 (32.8)	31 (47.0)	
Peritumoral enhancement				0.321
Absent	124 (80.0)	53 (79.1)	58 (87.9)	
Present	31 (20.0)	14 (20.9)	8 (12.1)	
Tumor margin				0.544
Absent	94 (60.6)	45 (67.2)	44 (66.7)	
Present	61 (39.4)	22 (32.8)	22 (33.3)	

Data are expressed as the mean ± SD or number (percentage).

AFP indicates alpha fetoprotein; ALT, alanine transferase; AST, aspartate transferase; FIB, fibrinogen; GGT, gamma-glutamyl transferase; INR, International Normalized Ratio; NLR, neutrophil-to-lymphocyte ratio; PLR, platelet-to-lymphocyte ratio; TB, total bilirubin.

peritumoral enhancement. When multiple lesions were present in the patient's liver, we evaluated the largest tumor. In cases where their assessments were discordant, a third senior radiologist was consulted to reach a consensus. Some example images and specific imaging semantic features are explained in Figure S1, Supplemental Digital Content 3, http://links.lww.com/JCAT/A28; eAppendix 3, Supplemental Digital Content 1, http://links.lww.com/JCAT/A26.

Image Segmentation

Physician A, who had 10 years of experience interpreting abdominal CT scans, used ITK-SNAP (version 3.6.0; http://www.itksnap.org) software to segment the images. The volume of interest (VOI)—comprising either the entire tumor or the residual liver excluding vessels or bile ducts—was delineated layer by layer on arterial, portal, and delayed phase images. In cases with multiple lesions, the largest lesion was chosen for segmentation. Throughout the delineation process, Physician A was blinded to the clinical data of the patients.

To extract the histologic features of the peritumoral images, the peritumoral VOIs were processed through the Python morphologic erosion and expansion algorithm, which automatically constricted the boundary of each lesion inwards by 5 mm and expanded it outwards by 3, 5, and 10 mm, respectively. To ensure the reproducibility of the radiomic features, physician A and physician B, each with 10 years of experience in reading abdominal CT images, independently repeated the segmentation procedure on the entire lesion for a randomly selected subset of 50 patients after a 2-week interval. The results of the repeat extraction were used to calculate the interclass correlation coefficient (ICC).

TABLE 2. Feature Selection Results From Univariate and Multivariate Analysis

	Univariable		Multivariable	
	OR (95% CI)	*P*	OR (95% CI)	*P*
Age				
< 43				
≥ 43	0.16 (0.04-0.60)	0.007	0.11 (0.02-0.55)	0.007
Cirrhosis				
Absent				
Present	0.32 (0.14-0.74)	0.008	0.29 (0.10-0.84)	0.022
TB (μmol/L)				
< 10.5				
≥ 10.5	0.39 (0.14-1.08)	0.069	0.32 (0.09-1.12)	0.075
FIB (g/L)				
< 2.22				
≥ 2.22	2.02 (1.04-3.91)	0.037	1.70 (0.74-3.90)	0.210
AFP (ng/ml)				
< 36.67				
≥ 36.67	2.19 (1.13-4.25)	0.020	2.63 (1.16-6.01)	0.021
Intratumor vascularity				
Absent				
Present	3.85 (1.85-8.02)	0.001	4.17 (1.68-10.38)	0.002
Intratumor necrosis				
Absent				
Present	1.89 (0.98-3.64)	0.058	1.07 (0.48-2.37)	0.865

AFP indicates alpha fetoprotein; FIB, Fibrinogen; TB, total bilirubin.

Feature Extraction Through Radiomics and Deep-Learning

Feature Extraction

Python version 3.8.4 (https://pypi.org/project/pyradiomics/) was used to extract radiomic features (eAppendix 4, Supplemental Digital Content 1, http://links.lww.com/JCAT/A26) for 3 image phases (arterial phase, A; portal phase, P; delayed phase, D) and 6 regions (tumor, residual liver, 5 mm-eroded and 3, 5, and 10 mm-extended peritumoral regions). To quantify the differences between different phases, we calculated the differences in the radiomic features between the A and P phases and between the D and P phases in the 6 regions (delta-radiomics). For each patient, 1539 imaging features were extracted. After removing features with a variance close to 0, the total number of features was reduced to 1512 in the tumor and residual liver regions and 1507 features in the other regions, resulting in a total of 45,290 features per patient. For more details, please refer to Table S1, Supplemental Digital Content 2, http://links.lww.com/JCAT/A27.

The settings for 3D ResNet-18 before feature extraction are described in eAppendix 5, Supplemental Digital Content 1, http://links.lww.com/JCAT/A26. In the deep-
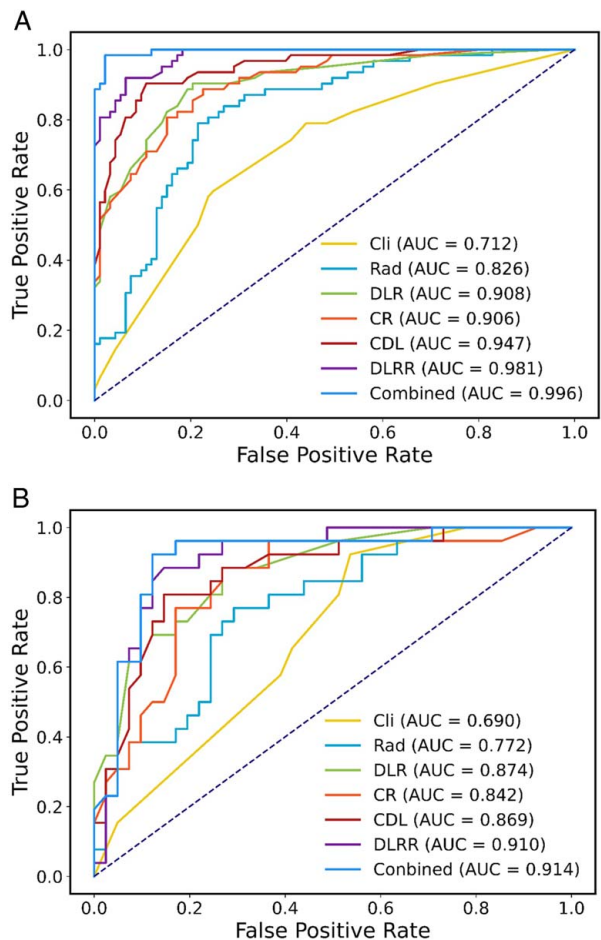

FIGURE 3. The ROC curves of each group of models. A, Performance of each group of models on the training set. B, Performance of each group of models on the internal validation set. ROC indicates receiver operating characteristic.

learning feature extraction process, the extraction area was the same as that used in the radiomic process. The number of features extracted from each region was 512, and a total of 9216 deep-learning features were extracted for each patient from 3 phases and 6 regions (Table S2, Supplemental Digital Content 2, http://links.lww.com/JCAT/A27), ensuring the consistency and comparability of the data.

Radiomic Features and Deep-Learning Feature Screening

To prevent overfitting, we used a 3-step feature selection process for feature screening. First, we selected features with an ICC > 0.8 and standardized these features through Z score normalization. Second, we used LASSO regression to select features with nonzero coefficients. Third, if the number of features exceeded one after LASSO selection, we further utilized recursive feature elimination (RFE) with a decision tree (DT) kernel to determine the optimal number of features at each stage. To obtain more representative features, each group of features underwent the above feature selection process. The specific feature selection process and results are shown in Figure S2, Supplemental Digital Content 3, http://links.lww.com/JCAT/A28, Tables S3, Supplemental Digital Content 2, http://links.lww.com/JCAT/A27. Finally, we calculated the Rad-score on the basis of the weighted regression coefficients of the radiomic features derived from LASSO.

In addition, we also conducted the above 3-step feature selection processes for features extracted through deep-learning. The specific feature selection results are shown in Table S4, Supplemental Digital Content 2, http://links.lww.com/JCAT/A27. We calculated the DL-score on the basis of the weighted regression coefficients of the deep-learning features derived from LASSO.

Model Building and Comparison

We built 7 models utilizing different features. The clinicoradiologic model (Cli) was constructed from clinical and imaging features resulting from univariate and multivariate analyses, the Rad-Score calculated from the radiomic features was used to construct a radiomic model (Rad), and the DL-score calculated from the deep-learning features was used to construct a deep-learning radiomic model (DLR). To identify potentially better prediction models, we constructed 4 integrated models based on the clinicoradiologic features, Rad-score and DL-score, including the clinicoradiologic and Rad-Score integrated model (CR), the clinicoradiologic and DL-score integrated model (CDL), the Rad-score and DL-score integrated model (DLRR), and a comprehensive model combining all features (Combined).

We constructed 7 models through 6 machine learning algorithms, including support vector machine (SVM), logistic regression (LR), random forest (RF), K-nearest neighbor (KNN), light gradient boosting machine (LightGBM), and Xtreme gradient boosting (XGBoost) algorithms. The most appropriate algorithm was selected on the basis of the characteristic data of the different groups to ensure the objectivity of the results. To improve the generalization ability of the models and better evaluate their performance with small sample sizes, we used 5-fold cross-validation for the model hyperparameter selection and model training process. Moreover, the process was performed on the training set only to avoid data leakage. In addition, we excluded models with an AUC > 0.1 between

TABLE 3. Six Best Models and Their Performance of Training and Internal Validation

Model		Train cohort				Internal validation			
		ACC	NPV	PPV	AUC (95% CI)	ACC	NPV	PPV	AUC (95% CI)
Cli	RF	0.69	0.581	0.763	0.712 (0.686–0.738)	0.612	0.654	0.585	0.690 (0.649–0.731)
Rad	LR	0.774	0.806	0.753	0.826 (0.790–0.862)	0.716	0.692	0.732	0.772 (0.718–0.826)
DLR	KNN	0.832	0.774	0.871	0.908 (0.855–0.961)	0.791	0.731	0.829	0.874 (0.795–0.953)
CR	LightGBM	0.832	0.806	0.849	0.906 (0.857–0.956)	0.776	0.769	0.78	0.842 (0.759–0.926)
CDL	LightGBM	0.884	0.855	0.903	0.947 (0.889–1.000)	0.791	0.808	0.78	0.869 (0.781–0.958)
DLRR	LightGBM	0.916	0.871	0.946	0.981 (0.917–1.000)	0.851	0.769	0.902	0.910 (0.815–1.000)
Combined	XGBoost	0.948	0.903	0.978	0.996 (0.931–1.000)	0.851	0.808	0.878	0.914 (0.817–1.000)

CDL indicates clinical imaging and DL-score combined model; Cli, clinical imaging model; CR, clinical imaging and Rad-score combined model; DLR, deep-learning model; DLRR, incorporates DLR and Rad model; Combined, a comprehensive model combining all features; Rad, radiomics model.

the training set and the internal validation set to avoid overfitting and underfitting. The best model with the highest AUC in the internal validation set was selected.

Model Validation and Clinical Application

To demonstrate the calibration ability and clinical applicability of the best model, we generated calibration curves and performed DCA on the training, internal validation and external validation sets.

To evaluate whether the best model can effectively predict progression-free survival (PFS) and overall survival (OS) for risk stratification, we utilized the maximum Youden index from the internal validation set as the optimal cutoff value for prediction outcomes in both the training and validation cohorts. Patients were categorized into low-risk and high-risk groups, and the 2-year PFS and 5-year OS rates were analyzed through Kaplan-Meier (K-M) survival curves.

Follow-Up

All patients were followed up regularly after discharge. The first follow-up was 1 month after the ablation procedure, during which the local therapeutic effect was evaluated. The other follow-ups were every 3 or 6 months after ablation. During follow-up visits, serum AFP levels, liver function tests, and CECT or MRI examinations were routinely performed, and Ultrasound angiography was used in selected cases if necessary. The starting point of this study was defined as the time at which the ablation procedure was performed, and the primary endpoint was ER. Recurrence was diagnosed based on the emergence of new intrahepatic lesions or metastases displaying imaging features characteristic of HCC or confirmed through histopathologic analysis. Curative ablation was defined as the absence of tumor necrosis enhancement on dynamic CECT, MRI, or CEUS. The last follow-up date for this study was March 31, 2024.

Statistical Analyses

Statistical analysis was performed through R software (version 4.3.0). In the training set, variables with more than 20% missing data were excluded; otherwise, multiple imputation algorithms were used to handle the missing data. To improve model interpretability, continuous variables were converted into binary variables through threshold values from receiver operating characteristic (ROC) curves. Categorical variables are presented as frequencies and percentages

and were analyzed through χ^2 tests or the Fisher exact tests. Variables showing a *p*-value of < 0.1 in the univariate analysis were included in the multivariate analysis for further variable selection. The models' predictive performance was
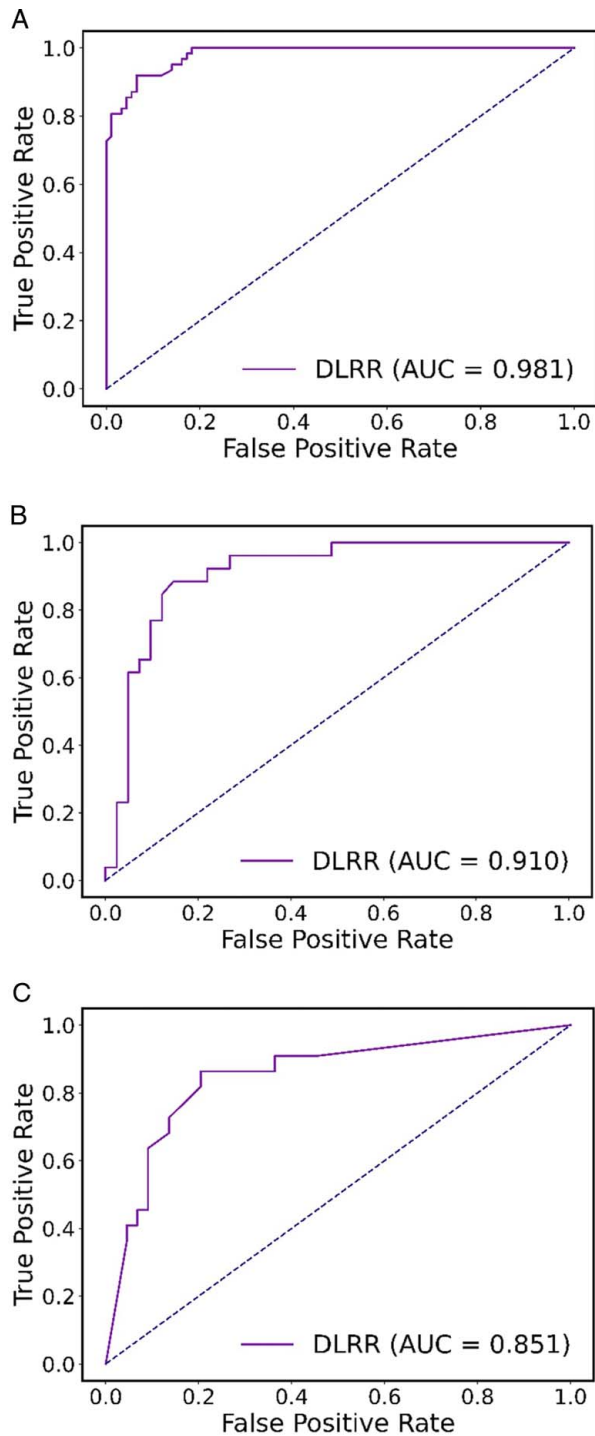

FIGURE 4. ROC curves for the training, internal validation and external validation sets. A, The ROC performance of the training set. B, The ROC performance of internal validation set. C, The ROC performance of external validation set. ROC indicates The receiver operating characteristic.

TABLE 4. Comparison Results of NRI and IDI for Different Models

	Model	Model	NRI (*P*)	IDI (*P*)
1	Cli	Rad	0.00249	0.03579
	Cli	DLR	< 0.0001	< 0.0001
	Rad	DLR	0.0453	0.03994
2	CR	CDL	0.23698	0.46977
	CR	DLRR	0.01142	0.0181
	CR	Combined	0.02276	0.00686
	CDL	DLRR	0.00077	6.00E-04
	CDL	Combined	0.0029	0.0082
	DLRR	Combined	0.56118	0.54389

CDL indicates clinical imaging and DL-score combined model; Cli, clinical imaging model; CR, clinical imaging and Rad-score combined model; DLR, deep-learning model; DLRR, incorporates DLR and Rad model, Combined, a comprehensive model combining all features; Rad, radiomics model.
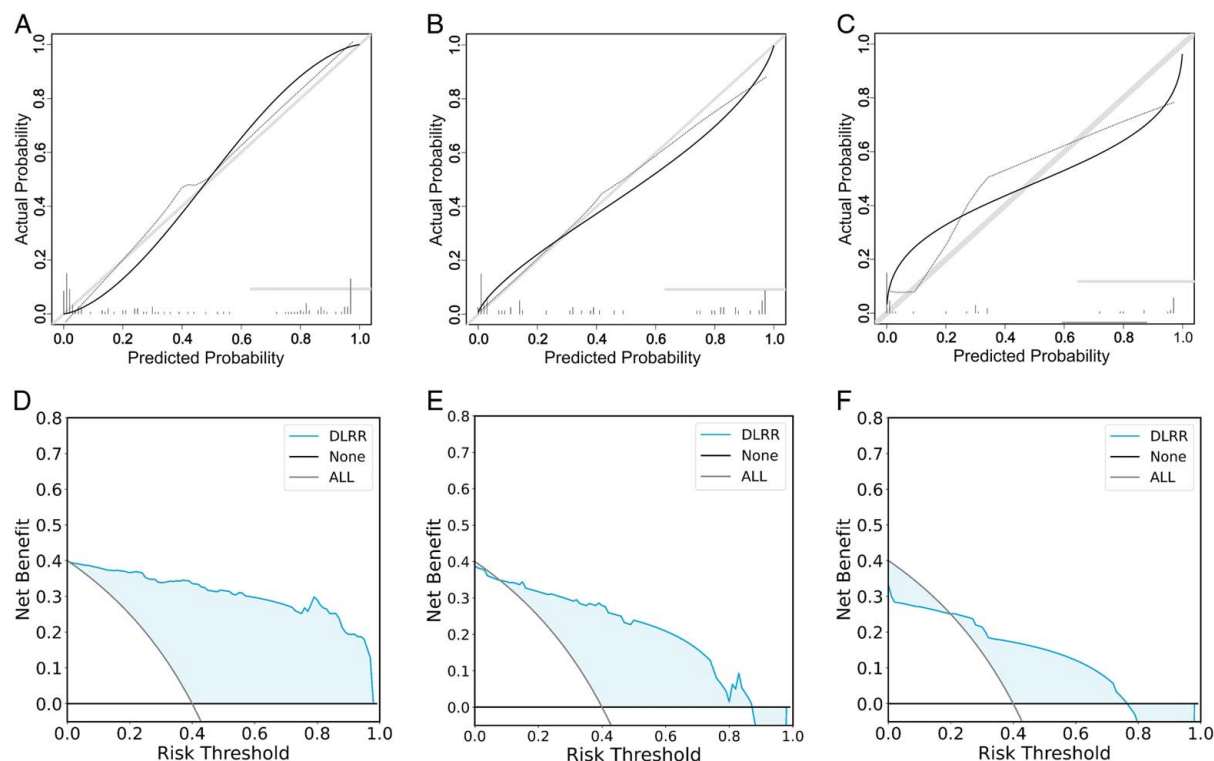
FIGURE 5. The calibration curve and DCA of DLRR. A–C, The calibration curve performance of DLRR on the training, internal validation, and external validation sets. D–F, The calibration curve performance of DLRR on the training, internal validation, and external validation sets.

assessed through the area under the curve (AUC), accuracy (ACC), positive predictive value (PPV), and negative predictive value (NPV). Model comparisons were made through the net reclassification improvement (NRI) and integrated discrimination improvement (IDI) indices. Calibration curves and decision curve analysis (DCA) were used to evaluate model calibration and clinical utility. K-M survival analysis was conducted, and log-rank tests were used for curve comparisons. A 2-tailed P-value of <0.05 was considered statistically significant.

RESULTS

Baseline Characteristics of the Study Cohorts

In Center 1, a total of 222 patients were included in the final analysis, with 155 patients assigned to the training cohort, and 67 patients were assigned to the internal validation cohort. In Centers 2 and 3, 66 patients were included in the independent external validation cohort. The clinicoradiologic characteristics of the patients are shown in Table 1. As of the final follow-up, the ER rate among the HCC patients was 38.2% (110/288).

The clinical information and imaging characteristics of the HCC patients in the training set, the internal validation set and the external validation set were not significantly different. The results of the univariate and multivariate analyses are presented in Table 2. Age ($P = 0.007$), AFP ($P = 0.021$), cirrhosis ($P = 0.022$), and the presence of intratumoral vessels($P = 0.002$) were found to be independent predictors of ER in patients with HCC.

Results of Radiomics and Deep-Learning Features Selection

Among the 3 imaging phases, 20 radiomic features (17 radiomic features and 3 delta-radiomic features) and 27 deep-learning features were found to be associated with ER in patients with HCC after curative ablation. The features and associated feature coefficients are shown in Table S5, Supplemental Digital Content 2, http://links.lww.com/JCAT/A27.

Results of Model Construction and Comparison

The Cli model was constructed from 4 clinicoradiologic features, while the Rad-Score calculated from 20 radiomic features was used to construct a radiomic model (Rad), and the DL-Score calculated from 27 deep-learning features was used to construct a deep-learning model (DLR). The ROC curves for each model are shown in Figure 3. According to a comparative analysis of the models, the DLR model outperformed the Cli and Rad models, with AUCs in the training set and internal validation set of 0.908 (0.855-0.961) and 0.874 (0.795-0.953), respectively. The comparison of AUCs between the DLR and Cli models showed a significant difference ($P < 0.0001$). In addition, a significant difference was also observed between the DLR and Rad models($P = 0.03994$).

With respect to the integrated models, Table 3 shows the best machine learning method corresponding to each model and their performance metrics. All methods and their performance metrics are shown in Table S6, Supplemental Digital Content 4, http://links.lww.com/JCAT/A29. The ROC curves in the training set and internal validation set
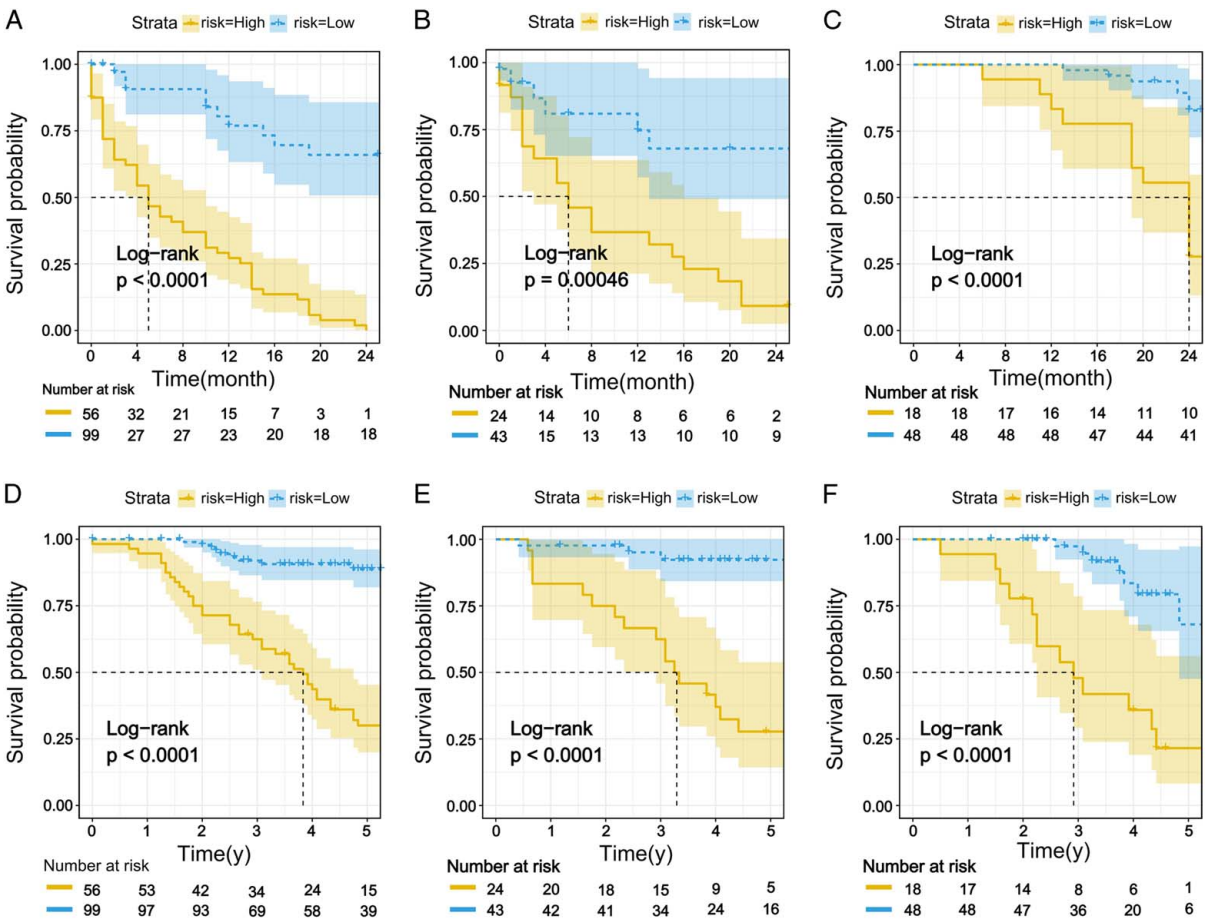
FIGURE 6. Internal training set, internal validation set, and external validation set exhibit Kaplan-Meier curves for 2-year progression-free survival (PFS) (A–C) and 5-year overall survival (OS) (D–F). The risk threshold dividing the high-risk group from the low-risk group is 0.72.

for each model are shown in Figure 3. Moreover, we determined the NRIs and IDIs for these models and found that the prediction effects of the combined model and the DLRR model were better than those of the other integrated models (Table 4). A detailed comparison of all models can be found in Table S7, Supplemental Digital Content 2, http://links.lww.com/JCAT/A27. The AUCs of the 2 models in the training set were 0.996 and 0.981, and those in the validation set were 0.914 and 0.910. Although the AUCs of the combined model in both the training set and the internal validation set were better than those of the DLRR model, the differences were not significant, indicating that the addition of clinicoradiologic features did not improve the predictive performance of the model. Considering that the DLRR model is more convenient than the combined model, as it can make predictions without the need to collect complicated clinical and imaging data, we chose the DLRR model as the optimal model for the following analyses.

Model Validation and Clinical Application

The DLRR model had an AUC of 0.981 in the training set, 0.910 in the internal validation set, and 0.851 in the external validation set (Fig. 4). Thus, the DLRR model has good generalizability. The DLRR model had good calibration and overall net benefits in both sets (Fig. 5).

Risk Stratification

The DLRR model could accurately stratify patients based on PFS and OS (both $P < 0.0001$) (Fig. 6). In addition, we demonstrated the process and corresponding results with 2 examples in Figures 7 and 8.

DISCUSSION

Accurately predicting ER after curative ablation of HCC is crucial for guiding precision therapy and improving patient prognosis.[26] Although the application of machine learning in radiomics has provided new predictive tools, the need to incorporate various clinical and biomarker data hinders its further promotion. Therefore, there is an urgent need to develop a more efficient and convenient model to predict ER in early-stage HCC patients. Our study collected preoperative CECT data and extracted features from 3 phases and 6 regions through DLRR methods. We constructed an integrated model based on associated features simultaneously. The results showed that the model had significant advantages in predicting ER, with AUCs of 0.981, 0.910, and 0.851 in the training, internal validation, and external validation sets, respectively. The K-M curves and corresponding cumulative risk curves clearly stratified both the training and validation sets.
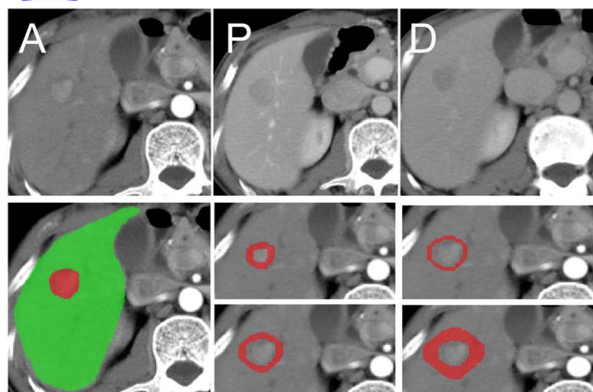
FIGURE 7. A typical case with CECT images and predicted results of response to ER by DLRR model. The red area represented the tumor tissue and peritumor area. The green area represented the residual liver tissue. A 42-year-old woman (patient 1) with ER. The probability of ER predicted by DLRR model was 0.836.
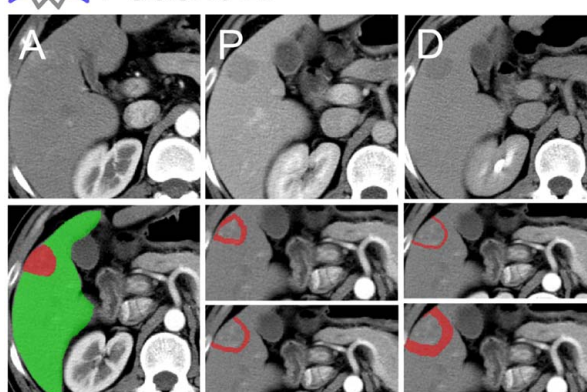
FIGURE 8. A typical case with CECT images and predicted results of response to ER by DLRR model. The red area represented the tumor tissue and peritumor area. The green area represented the residual liver tissue. A 32-year-old man (patient 2) without ER. The probability of ER predicted by DLRR model was 0.0148.

This study identified clinical and imaging features, and 4 variables were found to be independently associated with ER according to univariate and multivariate analyses, among which 3 factors, age, AFP, and cirrhosis, were consistent with the results of previous studies.[10,27,28] Bosi et al[7] confirmed that intratumoral vessels are associated with the growth of HCC. The AUC values in the training and validation sets of these 4 features were 0.712 and 0.690, respectively.

Radiomics was used to extract features from different phases and regions, and delta-radiomics was obtained by subtracting radiomic features from different phases. A total of 45290 features were extracted from each patient, and 17 radiomic features and 3 delta-radiomic features were ultimately obtained after 3 feature screening steps were applied. Features from the portal and delayed phases accounted for 82% (14/17) of the radiomic features and 74% (20/27) of the deep-learning features. This result was similar to that of the study by Yuan et al, who constructed a model for predicting recurrence based on 3-phase CT images and reported that the portal and delayed phase features accounted for 75% of the 20 related features.[19] The reason may be that although arterial phase images clearly reveal the abundant blood supply of early HCC tumors, the radiomic features of the portal and delayed phases could better reflect the microvascular structure and perfusion within the liver; this microlevel information is crucial for evaluating the prognostic effect of HCC ablation therapy.

Radiomic features in the surrounding tumor area play crucial roles in predicting early tumor recurrence.[29] In this study, 88% (15/17) of the features were in the peritumoral region, and 82% (14/17) of the features were within the 10 mm peritumoral range. The results demonstrated the importance of the peritumoral regional characteristics in predicting recurrence, similar to the findings of Zhou et al and Shi et al.[30,31] Therefore, in the absence of pathologic information, CECT images of the 10 mm peritumoral area can reveal features closely associated with ER.[32] Further analysis of the 17 radiomic features revealed that 30% (6/20) of the features were related to coarseness, a feature that describes the texture roughness of an image. To explore the distribution of this feature on the corresponding CT images, we used the "feature mapping" method to map the coarseness under different filters (eAppendix 6, Supplemental Digital Content 1, http://links.lww.com/JCAT/A26 for specific feature mapping steps). The results show that the coarseness features were mainly distributed in the boundary regions, with a large distribution in the 10 mm areas (Fig. S3, Supplemental Digital Content 3, http://links.lww.com/JCAT/A28; Fig. S4, Supplemental Digital Content 3, http://links.lww.com/JCAT/A28). Therefore, we suspected that this feature may be correlated with the heterogeneity of the peripheral microenvironment of the tumor and may influence tumor behavior and patient prognosis. We plan to perform further research in the future to confirm this conjecture.

CNNs are the primary deep-learning network for extracting features and can find deeper features than traditional radiomic methods can find, which reflect more important tumor information and make prediction easier.[33–35] In this study, 9216 deep-learning features were extracted for each patient from multiple phases and regions through 3D ResNet-18. After feature selection, 27 deep-learning features were ultimately used to construct the DL model. 89% (24/27) of the features were located in the peritumoral region, and 63% (17/27) of the features were within 10 mm of the area. Further analyses revealed that the AUCs of the predictive model in the training and validation groups were 0.908 and 0.874, respectively, which were significantly better than those of the clinicoradiologic model. This finding was consistent with the findings of Wu et al,[27] whose study reported a *C*-index of 0.695 (0.561-0.789) for predicting ER in the validation cohort, out-performing the clinical model. The superior performance of the model may be due to the extraction of features from multiple phases and regions, which may contain more prognostic information.

To increase the predictive efficiency of the model, Ma et al[36] integrated deep-learning and radiomic features, as well as integrated intratumoral and peritumoral regions, to provide more valuable information for predicting the therapeutic response of non-small cell lung cancer patients to chemoradiotherapy. Zhang et al[37] constructed a model based on 9 deep-learning features and 17 radiomic features to diagnose meningioma, achieving an AUC of 0.943 (0.873-1.000) in the test cohort, indicating potential clinical value for assisting doctors in preoperative tumor diagnosis. There have been no studies on the application of DLRR in predicting the prognosis of patients with HCC after ablation therapy. The AUCs of the DLRR of our model were 0.981 and 0.910 in the training and validation sets, respectively. Through NRI and IDI comparative analyses, we found that the DLRR model was significantly superior to the CR, DL, Rads, and Cli models. However, the combined model, which incorporates clinicoradiologic data into the DLRR model, was not significantly different from the DLRR model despite a slight increase in the AUC. This finding is consistent with the results of Ma et al.[11] The reason for this outcome may be that for early-stage HCC, the information provided by clinical and imaging data is limited due to the early-stage and small tumor size.

However, our study has several limitations. First, the retrospective nature of the study design may have inevitably introduced selection bias. Second, the relatively long study period may introduce bias related to treatment and imaging techniques. Nonetheless, the proposed model showed good prognostic performance in the training and test sets, suggesting that multiphase and multiregional CECT images had strong predictive value for the outcomes of ablation for HCC. Thirdly, the sample size of this study was relatively small, partly because the subjects of this study were early-stage HCC.

CONCLUSION

The DLRR model established in this study can noninvasively, efficiently, and conveniently predict ER after curative ablation in HCC patients through multiphase and multiregional CECT images. It can also stratify patients into risk subgroups based on PFS and OS rates. The successful establishment of this model provides new evidence for personalized treatment and offers close follow-up and additional treatment options for high-risk patients, thereby improving patient prognosis.

REFERENCES

1. Llovet JM, Kelley RK, Villanueva A, et al. Hepatocellular carcinoma. *Nat Rev Dis Primers*. 2021;7:6.
2. Villanueva A. Hepatocellular carcinoma. *N Engl J Med*. 2019; 380:1450–1462.
3. European Association for the Study of the Liver. EASL Clinical Practice Guidelines: Management of hepatocellular carcinoma. *J Hepatol*. 2018;69:182–236.
4. Doyle A, Gorgen A, Muaddi H, et al. Outcomes of radio-frequency ablation as first-line therapy for hepatocellular carcinoma less than 3 cm in potentially transplantable patients. *J Hepatol*. 2019;70:866–873.
5. Wang Z, Liu M, Zhang DZ, et al. Microwave ablation versus laparoscopic resection as first-line therapy for solitary 3-5-cm HCC. *Hepatology*. 2022;76:66–77.
6. Takayama T, Hasegawa K, Izumi N, et al. Surgery versus radiofrequency ablation for small hepatocellular carcinoma: a randomized controlled trial (SURF Trial). *Liver Cancer*. 2021; 11:209–218.
7. Bosi C, Rimini M, Casadei-Gardini A, et al. Understanding the causes of recurrent HCC after liver resection and radio-frequency ablation. *Expert Rev Anticancer Ther*. 2023;23: 503–515.
8. Wu Z, Zeng Y, Yuan Y, et al. Early recurrence of hepatocellular carcinoma in patients after ablation and resection: a propensity score analysis. *Am J Surg*. 2024;228: 94–101.
9. Nevola R, Ruocco R, Criscuolo L, et al. Predictors of early and late hepatocellular carcinoma recurrence. *World J Gastroenterol*. 2023;29:1243–1260.
10. Ma QP, He XL, Li K, et al. Dynamic contrast-enhanced ultrasound radiomics for hepatocellular carcinoma recurrence prediction after thermal ablation. *Mol Imaging Biol*. 2021;23: 572–585.
11. Yao LQ, Chen ZL, Feng ZH, et al. Clinical Features of Recurrence After Hepatic Resection for Early-Stage Hepatocellular Carcinoma and Long-Term Survival Outcomes of Patients with Recurrence: A Multi-institutional Analysis. *Ann Surg Oncol*. 2022;29:5206.
12. Hu C, Song Y, Zhang J, et al. Preoperative gadoxetic acid-enhanced MRI based nomogram improves prediction of early HCC recurrence after ablation therapy. *Front Oncol*. 2021;11: 649682.
13. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. 2012;48:441–446.
14. Ji GW, Zhu FP, Xu Q, et al. Radiomic features at contrast-enhanced CT predict recurrence in early stage hepatocellular carcinoma: a multi-institutional study. *Radiology*. 2020;294: 568–579.
15. Lee IC, Huang JY, Chen TC, et al. Evolutionary learning-derived clinical-radiomic models for predicting early recurrence of hepatocellular carcinoma after resection. *Liver Cancer*. 2021; 10:572–582.
16. Gao W, Wang W, Song D, et al. A predictive model integrating deep and radiomics features based on gadobenate dimeglumine-

enhanced MRI for postoperative early recurrence of hepatocellular carcinoma. *Radiol Med*. 2022;127:259–271.

17. Wang L, Song D, Wang W, et al. Data-driven assisted decision making for surgical procedure of hepatocellular carcinoma resection and prognostic prediction: development and validation of machine learning models. *Cancers (Basel)*. 2023;15:1784.

18. Qian GX, Xu ZL, Li YH, et al. Computed tomography-based radiomics to predict early recurrence of hepatocellular carcinoma post-hepatectomy in patients background on cirrhosis. *World J Gastroenterol*. 2024;30:2128–2142.

19. Yuan C, Wang Z, Gu D, et al. Prediction early recurrence of hepatocellular carcinoma eligible for curative ablation using a Radiomics nomogram. *Cancer Imaging*. 2019;19:21.

20. Peng W, Jiang X, Zhang W, et al. A radiomics-based model can predict recurrence-free survival of hepatocellular carcinoma after curative ablation. *Asian J Surg*. 2023;46:2689–2696.

21. Li Z, Wang Y, Yu J, et al. Deep Learning based Radiomics (DLR) and its usage in noninvasive IDH1 prediction for low grade glioma. *Sci Rep*. 2017;7:5467.

22. Wang F, Chen Q, Chen Y, et al. A novel multimodal deep learning model for preoperative prediction of microvascular invasion and outcome in hepatocellular carcinoma. *Eur J Surg Oncol*. 2023;49:156–164.

23. Wei Z, Liu H, Xv Y, et al. Development and validation of a CT-based deep learning radiomics nomogram to predict muscle invasion in bladder cancer. *Heliyon*. 2024;10:e24878.

24. Kocak B, Borgheresi A, Ponsiglione A, et al. Explanation and Elaboration with Examples for CLEAR (CLEAR-E3): an EuSoMII Radiomics Auditing Group Initiative. *Eur Radiol Exp*. 2024;8:72.

25. Heimbach JK, Kulik LM, Finn RS, et al. AASLD guidelines for the treatment of hepatocellular carcinoma. *Hepatology*. 2018;67:358–380.

26. Yang Y, Xin Y, Ye F, et al. Early recurrence after radiofrequency ablation for hepatocellular carcinoma: a multicenter retrospective study on definition, patterns and risk factors. *Int J Hyperthermia*. 2021;38:437–446.

27. Wu JP, Ding WZ, Wang YL, et al. Radiomics analysis of ultrasound to predict recurrence of hepatocellular carcinoma after microwave ablation. *Int J Hyperthermia*. 2022; 39:595–604.

28. Yang X, Yuan C, Zhang Y, et al. Predicting hepatocellular carcinoma early recurrence after ablation based on magnetic resonance imaging radiomics nomogram. *Medicine (Baltimore)*. 2022;101:e32584.

29. Shan QY, Hu HT, Feng ST, et al. CT-based peritumoral radiomics signatures to predict early recurrence in hepatocellular carcinoma after curative tumor resection or ablation. *Cancer Imaging*. 2019;19:11.

30. Zhou Z, Xia T, Zhang T, et al. Prediction of preoperative microvascular invasion by dynamic radiomic analysis based on contrast-enhanced computed tomography. *Abdom Radiol (NY)*. 2024;49:611–624.

31. Shi ZX, Li CF, Zhao LF, et al. Computed tomography radiomic features and clinical factors predicting the response to first transarterial chemoembolization in intermediate-stage hepatocellular carcinoma. *Hepatobiliary Pancreat Dis Int*. 2024;23:361–369.

32. Chen C, Han Q, Ren H, et al. Multiparametric MRI-based model for prediction of local progression of hepatocellular carcinoma after thermal ablation. *Cancer Med*. 2023;12: 17529–17540.

33. Poplin R, Varadarajan AV, Blumer K, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat Biomed Eng*. 2018;2:158–164.

34. Zhu Y, Man C, Gong L, et al. A deep learning radiomics model for preoperative grading in meningioma. *Eur J Radiol*. 2019; 116:128–134.

35. Liu Z, Jiang Z, Meng L, et al. Handcrafted and deep learning-based radiomic models can distinguish GBM from brain metastasis. *J Oncol*. 2021;2021:5518717.

36. Ma Y, Li Q. An integrated model combined intra- and peritumoral regions for predicting chemoradiation response of non small cell lung cancers based on radiomics and deep learning. *Cancer Radiother*. 2023;27:705–711.

37. Zhang J, Zhao Y, Lu Y, et al. Meningioma consistency assessment based on the fusion of deep learning features and radiomics features. *Eur J Radiol*. 2024;170:111250.