World Journal of Gastrointestinal Oncology

Submit a Manuscript: https://www.f6publishing.com

World J Gastrointest Oncol 2022 December 15; 14(12): 2380-2392

DOI: 10.4251/wjgo.v14.i12.2380

ISSN 1948-5204 (online)

ORIGINAL ARTICLE

Observational Study

Deep learning-based radiomics based on contrast-enhanced ultrasound predicts early recurrence and survival outcome in hepatocellular carcinoma

Zhe Huang, Zhu Shu, Rong-Hua Zhu, Jun-Yi Xin, Ling-Ling Wu, Han-Zhang Wang, Jun Chen, Zhi-Wei Zhang, Hong-Chang Luo, Kai-Yan Li

Specialty type: Oncology

Provenance and peer review: Unsolicited article; Externally peer reviewed.

Peer-review model: Single blind

Peer-review report's scientific quality classification

Grade A (Excellent): 0
Grade B (Very good): B, B
Grade C (Good): C
Grade D (Fair): 0
Grade E (Poor): 0

P-Reviewer: Gupta T, India; Qazi Arisar FA, Pakistan; Tsoulfas G, Greece

Received: September 15, 2022
Peer-review started: September 15, 2022
First decision: October 19, 2022
Revised: October 21, 2022
Accepted: November 21, 2022
Article in press: November 21, 2022
Published online: December 15, 2022

Zhe Huang, Zhu Shu, Jun-Yi Xin, Ling-Ling Wu, Hong-Chang Luo, Kai-Yan Li, Department of Medical Ultrasound, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, Hubei Province, China

Rong-Hua Zhu, Zhi-Wei Zhang, Institute of Hepato-pancreato-bililary Surgery, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, Hubei Province, China

Han-Zhang Wang, Jun Chen, PDx Advanced Applications, GE Healthcare, Shanghai 200020, China

Corresponding author: Kai-Yan Li, MD, Director, Doctor, Department of Medical Ultrasound, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, No. 95 Jiefang Avenue, Qiaokou District, Wuhan 430030, Hubei Province, China.
liky20006@126.com

Abstract

BACKGROUND

Hepatocellular carcinoma (HCC) is the most common primary liver malignancy.

AIM

To predict early recurrence (ER) and overall survival (OS) in patients with HCC after radical resection using deep learning-based radiomics (DLR).

METHODS

A total of 414 consecutive patients with HCC who underwent surgical resection with available preoperative grayscale and contrast-enhanced ultrasound images were enrolled. The clinical, DLR, and clinical + DLR models were then designed to predict ER and OS.

RESULTS

The DLR model for predicting ER showed satisfactory clinical benefits [area under the curve (AUC)] = 0.819 and 0.568 in the training and testing cohorts, respectively), similar to the clinical model (AUC = 0.580 and 0.520 in the training and testing cohorts, respectively; $P > 0.05$). The C-index of the clinical + DLR

model in the prediction of OS in the training and testing cohorts was 0.800 and 0.759, respectively. The clinical + DLR model and the DLR model outperformed the clinical model in the training and testing cohorts (P < 0.001 for all). We divided patients into four categories by dichotomizing predicted ER and OS. For patients in class 1 (high ER rate and low risk of OS), retreatment (microwave ablation) after recurrence was associated with improved survival (hazard ratio = 7.895, P = 0.005).

CONCLUSION
Compared to the clinical model, the clinical + DLR model significantly improves the accuracy of predicting OS in HCC patients after radical resection.

Key Words: Hepatocellular carcinoma; Deep learning; Overall survival; Early recurrence; Contrast-enhanced ultrasound

Core Tip: Multivariate Cox regression analysis confirmed that age [hazard ratio (HR) = 1.01], carbohydrate antigen 19-9 (HR = 0.60), tumor size (HR = 1.11), echogenicity (HR = 0.82), and deep learning-based radiomics (DLR, HR = 4.33) were independent predictors of survival outcome (P < 0.05 for all). The concordance index of the clinical + DLR model in the training and testing cohorts was 0.800 and 0.759, respectively. We divided patients into four categories by dichotomizing predicted early recurrence and survival outcome. We found that for patients with class 1 (high early recurrence rate and low risk of survival outcome), retreatment after recurrence was associated with improved survival.

Citation: Huang Z, Shu Z, Zhu RH, Xin JY, Wu LL, Wang HZ, Chen J, Zhang ZW, Luo HC, Li KY. Deep learning-based radiomics based on contrast-enhanced ultrasound predicts early recurrence and survival outcome in hepatocellular carcinoma. *World J Gastrointest Oncol* 2022; 14(12): 2380-2392
URL: https://www.wjgnet.com/1948-5204/full/v14/i12/2380.htm
DOI: https://dx.doi.org/10.4251/wjgo.v14.i12.2380

INTRODUCTION

Hepatocellular carcinoma (HCC) is the most common primary liver malignancy[1]. Surgical resection is considered the mainstream intervention for early HCC treatment[2], and its therapeutic effect has gradually improved in recent years. However, the postoperative recurrence rate of HCC remains as high as 60% at the 5-year follow-up[3], and the 5-year average survival rate is less than 32%[4]. Previous studies have reported that the prognosis of patients with early recurrence (ER) of HCC (within 1 year) after surgical resection is poorer than that of patients with late recurrence (> 1 year)[5]. Therefore, to develop future treatment strategies, there is an urgent need to improve the identification of patients at high risk of recurrence and poor prognosis; this may help identify those who may benefit from adjuvant systemic therapy.

Clinical biomarkers, such as tumor burden, associated with postoperative recurrence and outcomes have been identified[6]. However, the model based on these clinical biomarkers could not provide sufficient predictive power, and quantifiable measures and radiological information were not included in the model, which could provide essential information. Therefore, new representations of biomarker technology must be urgently explored to predict postoperative recurrence and patient outcomes more accurately.

Medical imaging is a potential method for HCC diagnosis[7]. A previous study has developed a model that used clinical and contrast-enhanced computed tomography-based radiographic features to accurately predict the ER of HCC after surgical resection[8]. Ultrasonography is widely used in HCC examination as it is cost-effective, widely available, and time-saving and provides real-time results. More importantly, contrast-enhanced ultrasonography (CEUS) can visualize the microcirculatory perfusion of HCC in real-time[9]. The microbubble contrast agent can be safely used in patients with decreased renal function[10].

Accurate prediction of early HCC recurrence and patient outcome is required to make clinical decisions before surgical resection. Traditional imaging features are relatively poor indicators of tumor heterogeneity (*e.g.*, microvascular invasion and recurrence), are poor predictors of clinical outcomes [11], and are highly subjective, difficult to quantify, and challenging to apply further. Radiomics analysis transforms raw images into countable quantitative features and interprets tumor pathophysiology[12]. Neural network mining to link these features to biological and clinical endpoints can

help develop models to predict patient outcomes, thereby improving prediction-based cancer management[13]. Deep learning-based radiomics (DLR) has been applied to patients with HCC and has achieved promising results in predicting microvascular invasion and response to transarterial chemoembolization[14,15].

This study aimed to predict ER and overall survival (OS) in patients with HCC after radical resection by establishing a DLR model using ultrasonographic and CEUS images.

MATERIALS AND METHODS

Patients

The institutional review board of our institution approved this retrospective study and waived the requirement to obtain written informed consent of all participants.

We retrospectively screened 5466 patients with HCC meeting the Milan criteria who underwent curative resection for HCC at our institution between January 2008 and December 2018. HCC was diagnosed based on pathological findings[16]. The exclusion criteria were as follows: (1) CEUS not performed; (2) recurrent lesions or a history of radiofrequency ablation, microwave ablation, or transcatheter arterial chemoembolization before CEUS; and (3) incomplete follow-up. A total of 5052 patients were excluded, leaving 414 in the final analysis (Supplementary Figure 1). Clinical features, including basic patient information, biological test results, and treatment-related information, were obtained from patient records.

CEUS has the following clinical indications: Patients at risk for HCC in whom ultrasound screening is positive for liver nodules according to current clinical practice standards as well as the World Federation for Ultrasound in Medicine and Biology (WFUMB), European Federation of Societies for Ultrasound in Medicine and Biology (EFSUMB), and CEUS Liver Imaging Reporting and Data System guidelines; focal liver lesions observed on single-phase CT or unenhanced MRI performed for other clinical purposes; indeterminate focal liver lesions observed multiphase contrast-enhanced CT or MRI; and definite HCC on CT or MRI images in reparation for or during tissue sampling, surgical resection, or percutaneous ablation treatment.

Ultrasonographic examination

CEUS was performed using a GE Logiq 9 scanner (GE Healthcare, Wauwatosa, WI, United States) with a 25-MHz frequency range transducer and a 3-5 L probe. Contrast agent (2.4 mL, SonoVue, Bracco, Milan, Italy) was injected intravenously, followed by flushing with 10 mL of 0.9% saline. Continuous observation of three-stage contrast enhancement was performed, including the arterial phase (0-30 s), portal phase (31-120 s), and late phase (121-360 s). CEUS inspections were recorded as video clips for analysis. For the CEUS image used for research, a frame of image with peak contrast intensity of the lesion was selected, that is, the frame of image with the maximum intensity was selected by quantitatively analyzing the enhanced intensity of the lesion in 0-360 s.

Two sonographers (one with more than 8 years of experience in CEUS and the other with more than 5 years of experience) who were blinded to the pathology results evaluated each lesion. In the event of a difference of opinion between the two readers, the final decision was made by a third blinded sonographer (with more than 20 years of CEUS experience). Tumor size, number, and satellite nodules on CEUS were evaluated. Nodules close to the primary tumor (< 3 cm) were designated as satellite nodules. During the late phase, the presence of perfusion defects surrounding HCC lesions was evaluated. If a nodule exhibited hypoechogenicity compared with the surrounding enhanced liver parenchyma, it was defined as the "presence of a perfusion defect of satellite nodules." The degree of enhancement of each lesion was compared with that of the surrounding normal liver tissue and classified as hyper-enhanced, iso-enhanced, or hypo-enhanced.

Follow-up protocol

The patients were followed up at 1 mo, 3 mo, 6 mo, 9 mo, and 12 mo after operation, and every 3-6 mo after 12 mo. At each follow-up, serum alpha-fetoprotein levels were measured and imaging (contrast-enhanced computed tomography, CEUS, or contrast-enhanced MRI) was performed. ER was defined as intrahepatic and/or extrahepatic recurrence of HCC within 1 year after resection. Given that an increasing serum alpha-fetoprotein level alone does not necessarily mean recurrence, recurrence was confirmed by radiological evidence of a new tumor. All patients were followed up until death, ER, or for at least 1 year after curative resection. OS was calculated as the time interval between the date of surgery and the date of death or last follow-up.

Model construction and validation

Image quantification (radiomics feature extraction and DLR feature extraction of grayscale image and CEUS image, which was a frame of image with peak contrast intensity of the lesion (Supplementary Figure 2), OS prognostic model construction and validation (including the development of

DLR score and OS prognostic model using clinical variables and DLR score, and validation of the prognostic models), and construction and validation of the model for ER prediction are shown in the Supplementary material. We also assessed the ability of the DLR model to improve the ability of three clinicians (with 11, 5, and 2 years of experience, respectively) to predict ER, with or without the assistance of the DLR model. To demonstrate the impact of the DLR model on clinician-individualized assessment performance, three clinicians independently reassessed each patient's ER status on the same day after accounting for the DLR model predictions.

Statistical analysis

Differences in the distribution of clinical variables of the training and testing cohorts were assessed by Fisher's exact test or the chi-square test for categorical data and the nonparametric Mann-Whitney test for continuous data. To evaluate the predictive performance of different models, we plotted the receiver operating characteristic (ROC) curve and calculated the area under the ROC curve (AUC). Accuracy, sensitivity, and specificity were calculated from the confusion matrix to quantitatively evaluate the predictive models. For prognostic models predicting ER, we used Kaplan-Meier analysis and the log-rank test to assess survival differences in prognostic clinical variables and radiological characteristics between the training and test sets. Interclass correlation coefficients (ICCs) were calculated for inter-observer and intra-observer agreement. A two-tailed *P*-value less than 0.05 was considered statistically significant. All statistical analyses and graphic production were completed with Python (version 3.8) and R (version 3.6.1).

RESULTS

Demographic characteristics

A total of 414 patients were included in this study, of which 289 and 125 were assigned to the training and testing cohorts, respectively. Mean age of the 414 patients was 53 (45-60) years, and 375 (90.6) were male. The demographic characteristics of patients in the training and testing cohorts were comparable (Table 1; $P > 0.05$). During a median follow-up period of 68 mo (range, 1-137 mo), 217 (52.4%) patients developed recurrence after curative surgical resection.

Prediction models for ER

We considered 414 grayscale ultrasonographic images and 414 peak contrast intensities of CEUS images. A total of 11270 radiological features were extracted from 414 patients. After univariate logistic regression selection, 898 features were retained. After the selection of the maximum relevance minimum redundancy, 8 features were selected as candidate features, and a preoperative ER prediction model was designed using L1-regularized logistic regression machine learning. Inter-observer ICC for measuring DLR features ranged from 0.633 to 0.989. Intra-observer ICC for measuring DLR features ranged from 0.689 to 0.927. The DLR model had an AUC of 0.819 in the training cohort with a 74% accuracy and an AUC of 0.568 in the test cohort with a 58% accuracy.

In this study, the presence of satellites was selected to construct a clinical model based on the constants and satellites (Table 2). The clinical model had an AUC of 0.58 in the training cohort with a 56% accuracy and an AUC of 0.52 in the test cohort with a 56% accuracy.

We constructed a clinical + DLR model, including the presence of satellites and radiomics/DL features. The AUC, accuracy, sensitivity, and specificity of the clinical + DLR model were 0.83, 73%, 71%, and 76%, respectively; the corresponding values in the testing cohort were 0.57, 59%, 62%, and 56%. The DLR model exhibited good classification performance based on ROC curves (Figure 1) and satisfactory clinical benefit (Figure 1), similar to those of the satellite lesion-based clinical model ($P > 0.05$). Prediction accuracy did not improve when clinical variables were combined with the DLR model (AUC: 0.830 for the clinical + DLR model *vs* 0.819 for the DLR model in the training cohort; AUC: 0.572 for the clinical + DLR model *vs* 0.568 for the DLR model in the testing cohort).

In univariate Fine-Gray regression analysis, satellite nodules, DLR model, and multiple lesions were significantly associated with ER. Using these variables, we performed a multivariate Fine-Gray competitive risk regression analysis. This analysis showed that the DLR model remained a strong independent predictor of ER after adjusting for clinical variables [odds ratio (OR) = 132.847, $P < 0.001$].

We found that human performance in predicting ER was significantly enhanced after integrating the DLR model. For clinician 1, the sensitivity increased significantly from 0.250 to 0.856 in the training cohort and from 0.230 to 0.812 in the testing cohort. Likewise, for clinicians 2 and 3, the sensitivity of both testing cohorts increased significantly (0.248 to 0.855 and 0.268 to 0.818 in the training cohort; 0.289 to 0.784 and 0.307 to 0.823 in the testing cohort). Scores were consistent across clinicians with a κ value of 0.76-0.89 and were significantly improved by the integrated DLR model, with a κ value of 0.93-0.97.

Prognostic model for OS

We obtained 11270 radiomics/DL features from patient images and selected them through univariate

Table 1 Patient characteristics

	Study population (*n* = 414)	Training cohort (*n* = 289)	Testing cohort (*n* = 125)	*P* value
Age, yr	53.00 (45.00- 60.00)	52.059 ± 12.190	53.216 ± 11.380	0.1988
Gender, *n* (%)				
Male	375 (90.6)	262 (90.7)	113 (90.4)	0.9196
BMI, kg/m^2	24.50 ± 4.20	24.10 ± 3.20	24.70 ± 3.40	0.1889
HBsAg-positive, *n* (%)	375 (90.6)	264 (91.3)	111 (88.8)	0.5274
AFP > 7 ng/mL, *n* (%)	292 (70.5)	205 (70.9)	87 (69.6)	0.8761
CEA > 5 ng/ml, *n* (%)	31 (7.5)	24 (8.3)	7 (5.6)	0.4494
CA125 > 40 ng/mL, *n* (%)	15 (3.6)	8 (2.8)	7 (5.6)	0.2588
CA199 > 34 ng/mL, *n* (%)	46 (11.1)	295 (10.0)	17 (13.6)	0.3738
WBC count, /μL	6262 ± 1985	6232 ± 1756	6354 ± 2125	0.5668
ALT, U/L	49 ± 36	48 ± 39	51 ± 41	0.1654
AST, U/L	51 ± 35	49 ± 36	53 ± 42	0.2358
Liver cirrhosis, *n* (%)	345 (83.3)	244 (84.4)	101 (80.8)	0.3630
Microvascular invasion, *n* (%)	312 (75.4)	211 (73.0)	101 (80.8)	0.0910
Tumor size, cm				
x	2.40 [1.70-3.68]	2.957 ± 1.850	3.120 ± 2.050	0.3689
y	2.00 [1.42-3.10]	2.480 ± 1.490	2.670 ± 1.960	0.4820
Gray-scale echogenicity				0.5954
Hyperechoic	46 (11.1)	35 (12.1)	11 (8.8)	
Medium	4 (1.0)	3 (1.0)	1 (0.8)	
Hypoechoic	364 (87.9)	251 (86.9)	113 (90.4)	
Arterial phase				0.4639
Hyperenhancement	403 (97.3)	283 (97.9)	120 (96.0)	
Isoenhancement	8 (1.9)	4 (1.4)	4 (3.2)	
Hypoenhancement	3 (0.7)	2 (0.7)	1 (0.8)	
Portal phase				0.6669
Hyperenhancement	15 (3.5)	12 (4.2)	3 (2.4)	
Isoenhancement	232 (56.0)	162 (56.1)	70 (56.0)	
Hypoenhancement	167 (40.3)	115 (39.8)	52 (41.6)	
Late phase				0.1300
Hyperenhancement	2 (0.5)	1 (0.3)	1 (0.8)	
Isoenhancement	79 (19.1)	46 (15.9)	33 (90.4)	
Hypoenhancement	333 (80.4)	232 (80.3)	101 (80.8)	
Enhancing capsules	45 (10.9)	36 (12.5)	9 (7.2)	0.1598
Unsmooth margins	97 (23.4)	64 (22.19)	33 (26.4)	0.4168
Retreatment after recurrence	168 (40.3)	118 (40.5)	50 (40.0)	0.9270

BMI: Body mass index; AFP: Alpha-fetoprotein; CEA: Carcinoembryonic antigen; CA125: Carbohydrate antigen 125; CA19-9: Carbohydrate antigen 19-9; BCLC: Barcelona-clinic liver cancer; WBC: White blood cell; HBsAg: Hepatitis B surface antigen; ALT: Alanine aminotransferase; AST: Aspartate aminotransferase.

Table 2 Univariate and multivariable analyses of early recurrence of hepatocellular carcinoma patients

	Univariate cox regression				Multivariate logistic regression			
	HR	[0.025	0.975]	*P*	HR	[0.025	0.975]	*P*
Age, yr	0.991	0.971	1.011	0.383				
Gender	1.005	0.463	2.181	0.990				
HBsAg-positive	1.138	0.472	2.740	0.773				
AFP	0.622	0.361	1.071	0.087	0.709	0.406	1.239	0.227
CEA	0.944	0.352	2.535	0.910				
CA125	1.811	0.424	7.752	0.423				
CA19-9	1.636	0.727	3.684	0.234				
ALT, U/L	1.248	0.697	2.321	0.267				
AST, U/L	1.566	0.397	2.108	0.675				
FIB-4 score	1.212	0.431	1.986	0.742				
Liver cirrhosis	1.142	0.506	2.121	0.657				
Tumor size x	1.000	0.882	1.143	0.951				
Tumor size y	0.988	0.856	1.142	0.873				
Gray-scale echogenicity	0.731	0.493	1.087	0.121				
Arterial phase enhancement	0.924	0.351	2.438	0.874				
Portal phase enhancement	1.321	0.832	2.100	0.238				
Portal phase enhancement	0.982	0.512	1.885	0.957				
Enhancing capsules	0.930	0.413	2.096	0.862				
Satellite nodules	4.843	1.917	12.244	0.001	4.194	1.368	12.871	0.012
Unsmooth margins	0.839	0.462	1.522	0.563				
Constant					0.990	0.707	1.387	0.953

Fibrosis-4 score: Age (years) × aspartate aminotransferase (U/L) / [platelet count (10^9/L) × alanine aminotransferase (U/L)]$^{1/2}$.
HR: Hazard ratio; AFP: Alpha-fetoprotein; CEA: Carcinoembryonic antigen; CA125: Carbohydrate antigen 125; CA19-9: Carbohydrate antigen 19-9; ALT: Alanine aminotransferase; AST: Aspartate aminotransferase; FIB-4 score: Fibrosis-4 score; HBsAg: Hepatitis B surface antigen.

Cox proportional hazards regression, with 50 features with a Harrell's concordance index (C-index) > 0.58. Multivariate Cox regression with L1 penalization calculates survival hazard values and builds high-risk and low-risk groups based on hazard values. The optimal stratification threshold for X-tile generation was 0.52. Kaplan-Meier curves showed significant differences between the low- and high-risk subgroups in the training and testing cohorts (Figure 2). Inter-observer ICC for measuring DLR features ranged from 0.611 to 0.976. Intra-observer ICC for measuring DLR features ranged from 0.699 to 0.912. The C-indices in the training and testing cohorts were 0.792 and 0.741, respectively. Calibration curves at 1, 3, and 5 years showed good agreement between DLR model estimates and actual observations in the training and testing cohorts.

In univariate analysis, nine significant factors, including five clinical variables (age, sex, carcinoembryonic antigen, carbohydrate antigen 125, and carbohydrate antigen 19-9), three semantic imaging features (tumor size x, tumor size y, and unsmooth margins), and DLR, were significantly associated with OS ($P < 0.05$) (Supplementary Table 1). Multivariate Cox regression analysis confirmed that age [hazard ratio (HR) = 1.01, 95% confidence interval (CI): 1.00-1.03, $P = 0.02$], carbohydrate antigen 19-9 (HR = 0.6, 95%CI: 0.04-1.03, $P = 0.007$), tumor size y (HR = 1.11, 95%CI: 1.03-1.19, $P = 0.001$), and DLR (HR = 4.33, 95%CI: 3.45-5.45, $P < 0.005$) were independent predictors of OS (Table 3). Based on the multivariate Cox regression analysis of assigned coefficients, these independent predictor variables were linearly combined to build a clinical and clinical + DLR model. Kaplan-Meier curves showed significant differences between the low- and high-risk subgroups in the training and testing cohorts (Figure 2). The C-indices of the clinical model in the training and testing cohorts were 0.566 and 0.565, respectively. The C-indices of the clinical + DLR model in the training and testing cohorts were 0.800 and 0.759, respectively. The clinical + DLR model and DLR model outperformed the clinical model in the training cohort ($P < 0.001$ and $P < 0.001$, respectively). Similar results were observed in the testing

Table 3 Univariate and multivariable analyses of overall survival of hepatocellular carcinoma patients

	Univariate cox regression				Multivariate logistic regression			
	HR	[0.025	0.975]	*P*	HR	[0.025	0.975]	*P*
Age, yr	1.020	1.000	1.030	0.010	1.02	1.00	1.03	0.01
Gender	1.570	1.030	2.390	0.030	1.42	0.89	2.26	0.14
HBsAg-positive	1.600	1.030	2.490	0.040	1.32	0.82	2.12	0.25
AFP	1.100	0.830	1.460	0.500				
CEA	0.760	0.460	1.270	0.300				
CA125	0.580	0.280	1.220	0.150				
CA19-9	0.630	0.420	0.950	0.030	0.65	0.41	1.03	0.07
ALT, U/L	1.121	0.453	1.976	0.430				
AST, U/L	1.342	0.876	2.014	0.540				
FIB-4 score	1.012	0.547	1.743	0.720				
Liver cirrhosis	1.112	0.563	1.956	0.550				
Tumor size x	1.080	1.010	1.160	0.040	0.96	0.85	1.09	0.56
Tumor size y	1.090	1.020	1.170	0.010	1.17	1.02	1.33	0.02
Gray-scale echogenicity	0.830	0.670	1.020	0.080	0.77	0.60	0.99	0.04
Arterial phase enhancement	0.680	0.420	1.110	0.130				
Portal phase enhancement	1.270	0.990	1.620	0.060	1.25	0.95	1.63	0.11
Portal phase Enhancement	1.130	0.810	1.580	0.460				
Enhancing capsules	1.110	0.720	1.710	0.630				
Satellite nodules	1.190	0.780	1.830	0.420				
Unsmooth margins	0.720	0.520	0.990	0.040	0.79	0.56	1.13	0.19
Early reoccurrence	1.290	0.990	1.680	0.060	1.25	0.93	1.67	0.14
Retreatment after recurrence	0.710	0.540	1.160	0.300				
DLR	3.240	2.670	3.930	< 0.005				

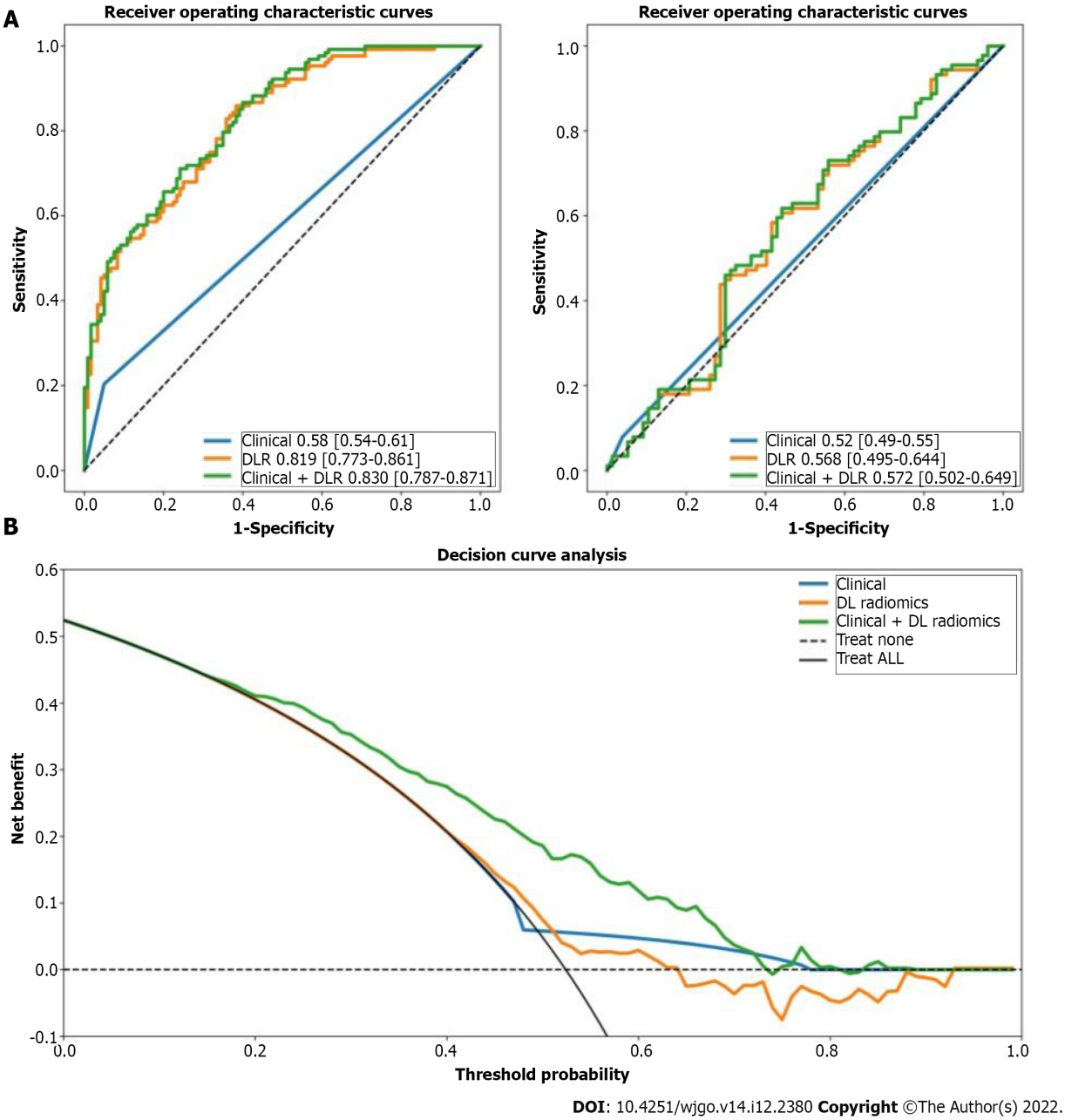
HR: Hazard ratio; AFP: Alpha-fetoprotein; CEA: Carcinoembryonic antigen; CA125: Carbohydrate antigen 125; CA19-9: Carbohydrate antigen 19-9; ALT: Alanine aminotransferase; AST: Aspartate aminotransferase; FIB-4 score: Fibrosis-4 score; DLR: Deep learning-based radiomics.

cohort ($P < 0.001$ and $P < 0.001$, respectively). The corresponding prediction error curves show that the prediction error of the clinical + DLR model was consistently lower than that of the clinical model. Similar results were obtained for the combined Brier scores in the training and testing cohorts. Finally, we further quantified the improvement in survival prediction accuracy between the clinical + DLR model and the clinical model. This resulted in a net reclassification improvement in survival of 0.234 (0.009 to 0.312; $P < 0.001$) and an OS net reclassification improvement of 0.176 (0.076 to 0.293; $P < 0.001$) in the testing cohort.

Histological features [degree of differentiation (HR = 1.76, 95%CI: 0.56-3.01, $P = 0.012$) and microvascular infiltration (HR = 2.25, 95%CI: 0.75-5.12, $P = 0.023$)] were independent predictors of OS (Supplementary Table 2). The clinical + DL and DLR models had the same performance with the histological features in the training cohort ($P = 0.157$ and $P = 0.566$, respectively). Similar results were observed in the testing cohort ($P = 0.225$ and $P = 0.648$, respectively).

Evaluation of the model for predicting OS and benefit of retreatment after recurrence

In addition to evaluating the accuracy of the model in predicting ER and OS, we further evaluated the correlation between retreatment (microwave ablation) after recurrence and OS in patients. We divided patients into four categories by dichotomizing predicted ER and OS. We found that for patients in class 1 (high ER rate and low risk of OS), retreatment after recurrence was associated with improved survival (HR = 7.895, $P = 0.005$). In contrast, for patients in class 2 (high ER rate and high risk of OS) (HR = 1.542, $P = 0.214$), class 3 (low ER rate and low risk of OS) (HR = 0.357, $P = 0.500$), and class 4 (low ER rate and



DOI: 10.4251/wjgo.v14.i12.2380

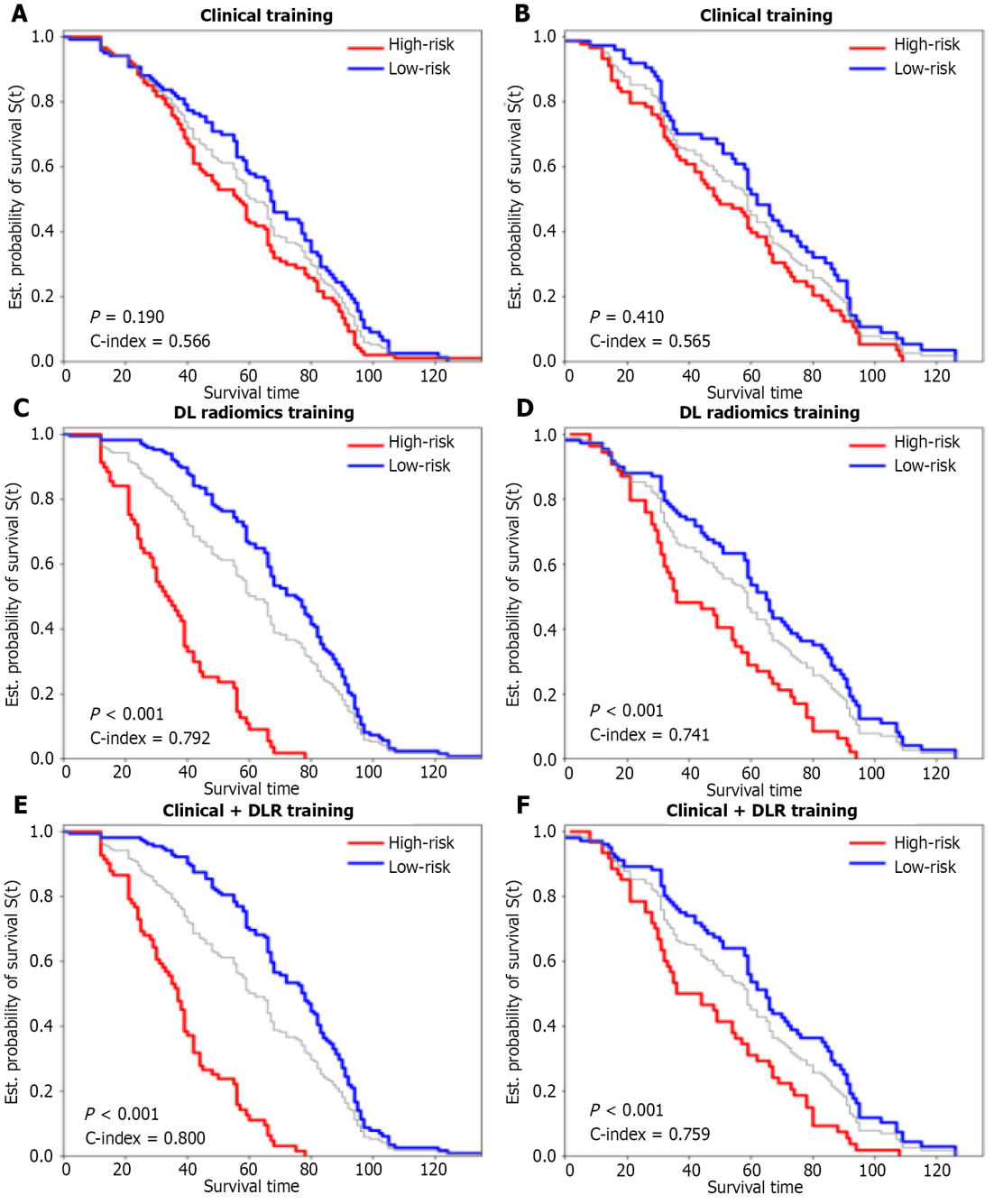
Figure 1 Receiver operating characteristic curves and decision curve analysis. A: Receiver operating characteristic curves of clinical, deep learning-based radiomics (DLR), and clinical + DLR models for predicting early recurrence in the training and testing cohorts; B: Decision curve analysis (DCA) of each model in predicting early recurrence. The vertical axis measures standardized net benefit. The horizontal axis shows the corresponding risk threshold. The DCA showed that if the threshold probability is between 0 and 1, using the DLR model derived in the present study to predict ER provided the same benefit as clinical model. ROC: Receiver operating characteristic; DCA: Decision curve analysis; DL: Deep learning; DLR: Deep learning-based radiomics.

high risk of OS) (HR = 1.416, P = 0.234), retreatment after recurrence did not affect survival (Figure 3).

DISCUSSION

This study aimed to develop and validate a predictive model for ER and OS in patients with early-stage HCC undergoing surgical resection. This model allows for better preoperative/pretreatment decision-making as to the best possible treatment options and timing. We transformed radiomics/DL signatures into quantitative radiomics/DL signatures and constructed a DLR model with a better ability to predict patient OS preoperatively than clinical models alone. This model may guide individualized treatment and survival monitoring.

Early-stage HCC still has a high recurrence rate after radical surgery. In our study, 52.4% of postoperative patients developed ER. Early HCC has high heterogeneity and different prognoses, which

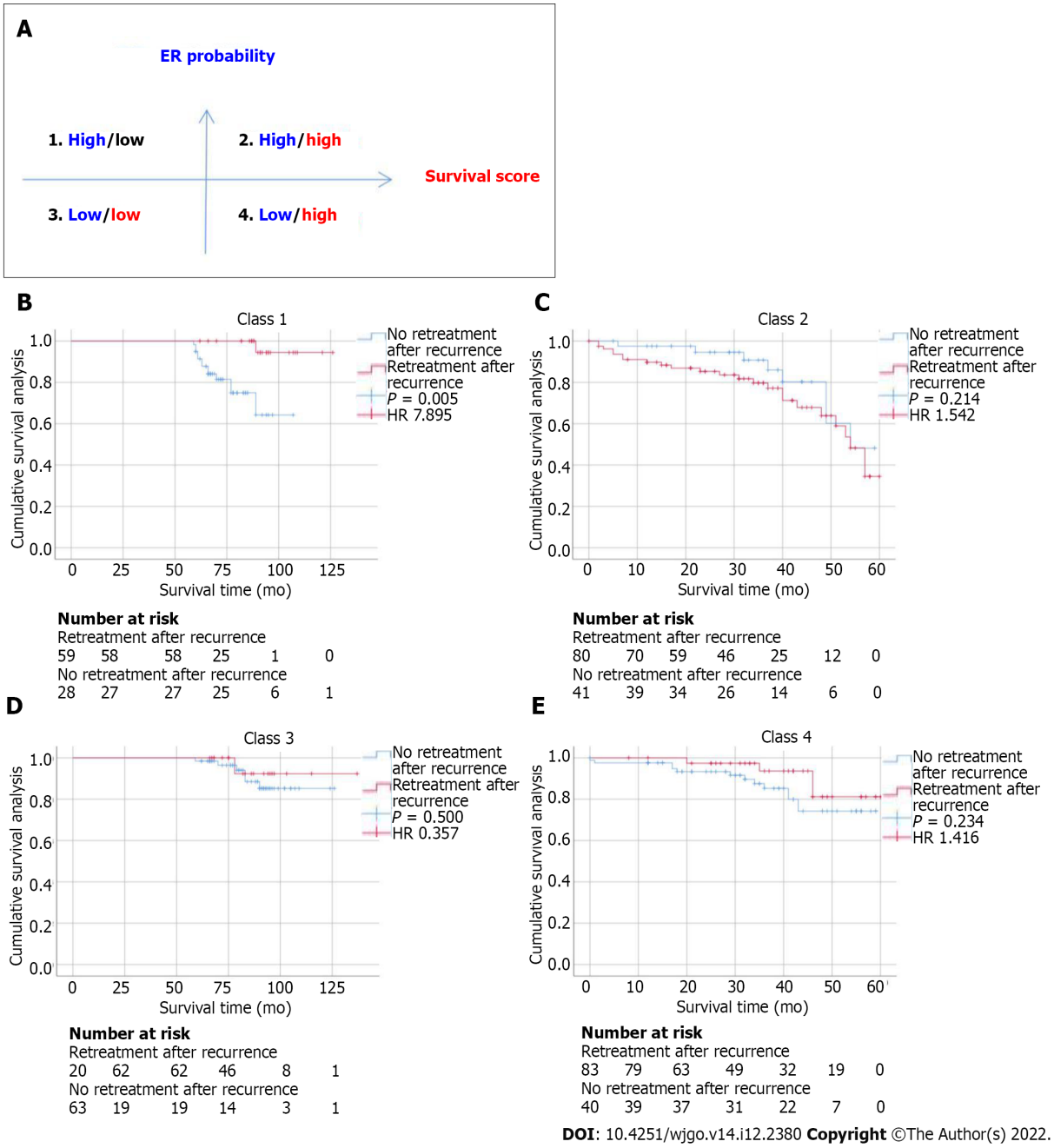


DOI: 10.4251/wjgo.v14.i12.2380

Figure 2 Kaplan-Meier curves of overall survival stratified by high and low risk for clinical, deep learning-based radiomics, and clinical + deep learning-based radiomics models. A and B: Clinical training; C and D: Deep learning-based radiomics testing; E and F: Clinical + DLR. DLR: Deep learning-based radiomics.

should be determined early and accurately. Gene signatures have been widely used in tumor identification but are rarely used in clinical applications because of their high cost and time consumption. Considering the high recurrence rate of HCC after radical resection, including disseminated or recurrent disease, early prediction of ER is critical for improving individualized treatment. As a routine examination, ultrasonography has a high potential for further investigation of ER-predictive radiological features. With the development of machine learning technology, a large amount of quantitative radiological data has been used to construct more predictive models than those developed by semantic radiological features. Our study found that the DLR score has the same or a higher ability to predict ER than satellite nodules. The DLR score can predict the patient's ER before surgery, helping guide treatment choices.

In this study, we developed an ER-related DLR model and evaluated its role in predicting OS. Furthermore, we combined clinical and DLR features to predict OS. In the multivariate analysis, we found that age was a significant risk factor for OS in patients with HCC, consistent with the results of



DOI: 10.4251/wjgo.v14.i12.2380

Figure 3 Relationship between the deep learning-based radiomics model and benefit from retreatment after recurrence in matched patients. A: Four different risk classes were defined by early recurrence and overall survival predicted by the deep learning-based radiomics model; B-E: Kaplan-Meier curves of disease-free survival for patients who were stratified according to receipt of retreatment after recurrence. HR: Hazard ratio.

other studies[8]. In addition, we found carbohydrate antigen 19-9, tumor size, and echogenicity were important risk factors for OS. However, the clinical + DLR and DLR models made a more dominant contribution in predicting OS than these clinical variables.

Re-surgical resection is considered a treatment for HCC recurrence. A key treatment issue is how to identify patients who may benefit from retreatment for HCC recurrence. However, given the costs associated with treatment, surgical trauma, and modest survival benefits, the optimal criteria for selecting candidates for retreatment for HCC recurrence remain unclear. The model developed in this study can help identify such patients. By combining information on ER risk and survival, our model can identify patients in class 1 who are more likely to benefit from re-surgical resection treatment.

HCC is a tumor with rich blood supply in which tortuous and dilated new vessels are continuously generated. In our study, the frame with the highest peak intensity in the arterial phase of CEUS was used, reflecting the density of neovascularization in the tumor. Studies have shown that the peak intensity in the recurrence group is lower than that in the non-recurrence group and that peak intensity is a risk factor for HCC recurrence[17].

Liu *et al*[18] analyzes CEUS images based on DLR to predict progression-free survival after radiofrequency ablation and surgical resection and optimize radiofrequency ablation and surgical resection treatment options for patients with HCC. Our study was also based on DLR analysis of CEUS images; the difference is that our study aimed to predict the ER and OS of patients with HCC after surgical resection and provide guidance for retreatment after recurrence.

Our study has two significant limitations. First, this is a single-center study, and a multicenter prospective study with a larger patient population is needed to further validate the performance of our model. Second, regions of interest were segmented manually and have not been fully automated. Finally, washout is also an important aspect of the assessment of HCC, and other images of CEUS need to be fully studied in the future.

CONCLUSION

The DLR model has the same satisfactory clinical benefit for predicting ER as the clinical model. Compared with the clinical model, the clinical + DLR model and the DLR model significantly improve the accuracy of predicting OS in HCC patients after radical resection.

ARTICLE HIGHLIGHTS

Research background

Hepatocellular carcinoma (HCC) is the most common primary liver malignancy.

Research motivation

To develop future treatment strategies, there is an urgent need to improve the identification of patients at high risk of recurrence and poor prognosis; this may help identify those who may benefit from adjuvant systemic therapy.

Research objectives

To predict early recurrence (ER) and overall survival (OS) in patients with HCC after radical resection using deep learning (DL)-based radiomics.

Research methods

A total of 414 consecutive patients with HCC who underwent surgical resection with available preoperative grayscale and contrast-enhanced ultrasound images were enrolled. The clinical, DLR, and clinical + DLR model were then designed to predict ER and OS.

Research results

The DLR model for predicting ER showed satisfactory clinical benefits [area under the curve (AUC)] = 0.819 and 0.568 in the training and testing cohorts, respectively), similar to the clinical model (AUC 0.580 and 0.520 in the training and testing cohorts, respectively; $P > 0.05$). The C-indices of the clinical + DLR model in prediction of OS in the training and testing cohorts were 0.800 and 0.759, respectively. The clinical + DLR model and the DLR model outperformed the clinical model in the training and testing cohorts ($P < 0.001$ for all). We divided patients into four categories by dichotomizing predicted ER and OS. For patients in class 1 (high ER rate and low risk of OS), retreatment (microwave ablation) after recurrence was associated with improved survival (hazard ratio = 7.895, $P = 0.005$).

Research conclusions

As compared to the clinical model, the clinical + DLR model significantly improves the accuracy of predicting OS in patients with HCC after radical resection.

Research perspectives

As compared to the clinical model, the clinical + DLR model significantly improves the accuracy of predicting OS in patients with HCC after radical resection.

FOOTNOTES

Author contributions: Huang Z contributed to study concept and design, data acquisition, analysis, and interpretation, and manuscript drafting; Wang HZ and Chen J contributed to the implementation and analysis of all machine learning methods; Huang Z, Zhu S, Wu XB, and Zhu RH contributed to image interpretation and segmentation; Zhu S, Wu XB, Zhang ZW, and Li KY contributed to data acquisition; Huang Z, Li KY, Wang HZ, and

Zhu RH contributed to study concept and design, critical revision of the manuscript for important intellectual content, and study supervision.

Institutional review board statement: The study was reviewed and approved by the Ethics Committee of Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology (Approval No. TJ-IRB20190401).

Informed consent statement: The Ethics Committee of Tongji Hospital waived the requirement to obtain written informed consent of all participants.

Conflict-of-interest statement: There are no conflicts of interest to report.

Data sharing statement: All data generated or analyzed during this study are included in this article and/or its supplementary material files. Further inquiries can be directed to the corresponding author.

STROBE statement: The authors have read the STROBE Statement—checklist of items, and the manuscript was prepared and revised according to the STROBE Statement—checklist of items.

Country/Territory of origin: China

ORCID number: Rong-Hua Zhu 0000-0002-8588-7493; Kai-Yan Li 0000-0003-3332-6325.

S-Editor: Chen YL
L-Editor: Wang TQ
P-Editor: Chen YL

REFERENCES

1 **Omata M**, Cheng AL, Kokudo N, Kudo M, Lee JM, Jia J, Tateishi R, Han KH, Chawla YK, Shiina S, Jafri W, Payawal DA, Ohki T, Ogasawara S, Chen PJ, Lesmana CRA, Lesmana LA, Gani RA, Obi S, Dokmeci AK, Sarin SK. Asia-Pacific clinical practice guidelines on the management of hepatocellular carcinoma: a 2017 update. *Hepatol Int* 2017; **11**: 317-370 [PMID: 28620797 DOI: 10.1007/s12072-017-9799-9]

2 **Hobeika C**, Nault JC, Barbier L, Schwarz L, Lim C, Laurent A, Gay S, Salamé E, Scatton O, Soubrane O, Cauchy F. Influence of surgical approach and quality of resection on the probability of cure for early-stage HCC occurring in cirrhosis. *JHEP Rep* 2020; **2**: 100153 [PMID: 32995713 DOI: 10.1016/j.jhepr.2020.100153]

3 **Lim KC**, Chow PK, Allen JC, Siddiqui FJ, Chan ES, Tan SB. Systematic review of outcomes of liver resection for early hepatocellular carcinoma within the Milan criteria. *Br J Surg* 2012; **99**: 1622-1629 [PMID: 23023956 DOI: 10.1002/bjs.8915]

4 **El-Serag HB**. Hepatocellular carcinoma. *N Engl J Med* 2011; **365**: 1118-1127 [PMID: 21992124 DOI: 10.1056/NEJMra1001683]

5 **Poon RT**, Fan ST, Ng IO, Lo CM, Liu CL, Wong J. Different risk factors and prognosis for early and late intrahepatic recurrence after resection of hepatocellular carcinoma. *Cancer* 2000; **89**: 500-507 [PMID: 10931448]

6 **Villanueva A**. Hepatocellular Carcinoma. *N Engl J Med* 2019; **380**: 1450-1462 [PMID: 30970190 DOI: 10.1056/NEJMra1713263]

7 **Llovet JM**, Brú C, Bruix J. Prognosis of hepatocellular carcinoma: the BCLC staging classification. *Semin Liver Dis* 1999; **19**: 329-338 [PMID: 10518312 DOI: 10.1055/s-2007-1007122]

8 **Lee IC**, Huang JY, Chen TC, Yen CH, Chiu NC, Hwang HE, Huang JG, Liu CA, Chau GY, Lee RC, Hung YP, Chao Y, Ho SY, Huang YH. Evolutionary Learning-Derived Clinical-Radiomic Models for Predicting Early Recurrence of Hepatocellular Carcinoma after Resection. *Liver Cancer* 2021; **10**: 572-582 [PMID: 34950180 DOI: 10.1159/000518728]

9 **Alzaraa A**, Gravante G, Chung WY, Al-Leswas D, Morgan B, Dennison A, Lloyd D. Contrast-enhanced ultrasound in the preoperative, intraoperative and postoperative assessment of liver lesions. *Hepatol Res* 2013; **43**: 809-819 [PMID: 23745715 DOI: 10.1111/hepr.12044]

10 **Wilson SR**, Feinstein SB. Introduction: 4th Guidelines and Good Clinical Practice Recommendations for Contrast Enhanced Ultrasound (CEUS) in the Liver-Update 2020 WFUMB in Cooperation with EFSUMB, AFSUMB, AIUM and FLAUS. *Ultrasound Med Biol* 2020; **46**: 3483-3484 [PMID: 32888748 DOI: 10.1016/j.ultrasmedbio.2020.08.015]

11 **Renzulli M**, Brocchi S, Cucchetti A, Mazzotti F, Mosconi C, Sportoletti C, Brandi G, Pinna AD, Golfieri R. Can Current Preoperative Imaging Be Used to Detect Microvascular Invasion of Hepatocellular Carcinoma? *Radiology* 2016; **279**: 432-442 [PMID: 26653683 DOI: 10.1148/radiol.2015150998]

12 **Bi WL**, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, Allison T, Arnaout O, Abbosh C, Dunn IF, Mak RH, Tamimi RM, Tempany CM, Swanton C, Hoffmann U, Schwartz LH, Gillies RJ, Huang RY, Aerts HJWL. Artificial intelligence in cancer imaging: Clinical challenges and applications. *CA Cancer J Clin* 2019; **69**: 127-157 [PMID:

30720861 DOI: 10.3322/caac.21552]
13 **Avanzo M**, Wei L, Stancanello J, Vallières M, Rao A, Morin O, Mattonen SA, El Naqa I. Machine and deep learning methods for radiomics. *Med Phys* 2020; **47**: e185-e202 [PMID: 32418336 DOI: 10.1002/mp.13678]
14 **Xu X**, Zhang HL, Liu QP, Sun SW, Zhang J, Zhu FP, Yang G, Yan X, Zhang YD, Liu XS. Radiomic analysis of contrast-enhanced CT predicts microvascular invasion and outcome in hepatocellular carcinoma. *J Hepatol* 2019; **70**: 1133-1144 [PMID: 30876945 DOI: 10.1016/j.jhep.2019.02.023]
15 **Chen M**, Cao J, Hu J, Topatana W, Li S, Juengpanich S, Lin J, Tong C, Shen J, Zhang B, Wu J, Pocha C, Kudo M, Amedei A, Trevisani F, Sung PS, Zaydfudim VM, Kanda T, Cai X. Clinical-Radiomic Analysis for Pretreatment Prediction of Objective Response to First Transarterial Chemoembolization in Hepatocellular Carcinoma. *Liver Cancer* 2021; **10**: 38-51 [PMID: 33708638 DOI: 10.1159/000512028]
16 **European Association for the Study of the Liver**. EASL Clinical Practice Guidelines: Management of hepatocellular carcinoma. *J Hepatol* 2018; **69**: 182-236 [PMID: 29628281 DOI: 10.1016/j.jhep.2018.03.019]
17 **Xuan Z**, Wu N, Li C, Liu Y. Application of contrast-enhanced ultrasound in the pathological grading and prognosis prediction of hepatocellular carcinoma. *Transl Cancer Res* 2021; **10**: 4106-4115 [PMID: 35116708 DOI: 10.21037/tcr-21-1264]
18 **Liu F**, Liu D, Wang K, Xie X, Su L, Kuang M, Huang G, Peng B, Wang Y, Lin M, Tian J. Deep Learning Radiomics Based on Contrast-Enhanced Ultrasound Might Optimize Curative Treatments for Very-Early or Early-Stage Hepatocellular Carcinoma Patients. *Liver Cancer* 2020; **9**: 397-413 [PMID: 32999867 DOI: 10.1159/000505694]

Published by **Baishideng Publishing Group Inc**
7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA
Telephone: +1-925-3991568
E-mail: bpgoffice@wjgnet.com
Help Desk: https://www.f6publishing.com/helpdesk
https://www.wjgnet.com