

Radiomic Features at Contrast-enhanced CT Predict Recurrence in Early Stage Hepatocellular Carcinoma: A Multi-Institutional Study

Gu-Wei Ji, MD, PhD • Fei-Peng Zhu, MD, PhD • Qing Xu, MD, PhD • Ke Wang, MD, PhD •
Ming-Yu Wu, MD • Wei-Wei Tang, MD • Xiang-Cheng Li, MD, PhD • Xue-Hao Wang, MD

From the Hepatobiliary Center (G.W.J., K.W., X.C.L., X.H.W.) and Department of Radiology (F.P.Z., Q.X.), The First Affiliated Hospital of Nanjing Medical University, 300 Guangzhou Road, Nanjing, P.R. China; Key Laboratory of Liver Transplantation, Chinese Academy of Medical Sciences, Nanjing, P.R. China (G.W.J., K.W., X.C.L., X.H.W.); NHC Key Laboratory of Living Donor Liver Transplantation, Nanjing Medical University, Nanjing, P.R. China (G.W.J., K.W., X.C.L., X.H.W.); Department of Hepatobiliary Surgery, Wuxi People's Hospital, Wuxi, P.R. China (M.Y.W.); and Department of General Surgery, Nanjing First Hospital, Nanjing, P.R. China (W.W.T.). Received July 3, 2019; revision requested August 28; revision received October 31; accepted November 7. **Address correspondence to** X.H.W. (e-mail: wangxh@njmu.edu.cn).

Supported by the Key Program of the National Natural Science Foundation of China (grant 31930020), the National Natural Science Foundation of China (grants 81530048, 81470901, and 81670570), and the Jiangsu Provincial Key Research and Development Program (grant BE2016789).

Conflicts of interest are listed at the end of this article.

Radiology 2020; 00:1–13 • <https://doi.org/10.1148/radiol.2020191470> • Content codes: **CT** **GI** **IN**

Background: Early stage hepatocellular carcinoma (HCC) is the ideal candidate for resection in patients with preserved liver function; however, cancer will recur in half of these patients and no reliable prognostic tool has been established.

Purpose: To investigate the effectiveness of radiomic features in predicting tumor recurrence after resection of early stage HCC.

Materials and Methods: In total, 295 patients (median age, 58 years; interquartile range, 50–65 years; 221 men) who underwent contrast material-enhanced CT and curative resection for early stage HCC that met the Milan criteria between February 2009 and December 2016 were retrospectively recruited from three independent institutions. Follow-up consisted of serum α -fetoprotein level, liver function tests, and dynamic imaging examinations every 3 months during the first 2 years and then every 6 months thereafter. In the development cohort of 177 patients from institution 1, recurrence-related radiomic features were computationally extracted from the tumor and its periphery and a radiomics signature was built with least absolute shrinkage and selection operator regression. Two models, one integrating preoperative and one integrating pre- and postoperative variables, were created by using multivariable Cox regression analysis. An independent external cohort of 118 patients from institutions 2 and 3 was used to validate the proposed models.

Results: The preoperative model integrated radiomics signature with serum α -fetoprotein level and tumor number; the postoperative model incorporated microvascular invasion and satellite nodules into the above-mentioned predictors. In both study cohorts, two radiomics-based models provided better predictive performance (concordance index ≥ 0.77 , $P < .05$ for all), lower prediction error (integrated Brier score ≤ 0.14), and larger net benefits, as determined by means of decision curve analysis, than rival models without radiomics and widely adopted staging systems. The radiomics-based models gave three risk strata with high, intermediate, or low risk of recurrence and distinct profiles of recurrent tumor number.

Conclusion: The proposed radiomics models with pre- and postresection features helped predict tumor recurrence for early stage hepatocellular carcinoma.

© RSNA, 2020

Online supplemental material is available for this article.

Hepatocellular carcinoma (HCC) is the sixth leading type of cancer and the second most fatal tumor worldwide (1). For patients with early stage HCC as defined by the Milan criteria (solitary nodule ≤ 5 cm or as many as three nodules ≤ 3 cm, without macrovascular invasion and extrahepatic spread), both liver resection and liver transplant are the mainstay curative options (1,2). Although liver transplant offers definite advantages of extirpating both the tumor and the diseased liver, demand for organs far exceeds supply. Therefore, liver resection is accepted as the first-line treatment option for patients with early stage HCC and preserved liver function, whereas liver transplant is the recommended treatment for patients with decompensated cirrhosis (3). Unfortunately, HCC recurrence, including true recurrence

by means of tumor dissemination and development of de novo tumors in the cirrhotic liver, occurs in 50%–60% of these patients at 5 years (4,5).

Currently, HCC staging systems (eg, Barcelona Clinic Liver Cancer, Hong Kong Liver Cancer, Cancer of the Liver Italian Program, and TNM systems) occupy the central role in prognosis and therefore treatment allocation (1). Accurate risk prediction allows optimal surveillance, prevention, and management strategies for tumor recurrence; however, these systems are inadequate for predicting recurrence, and none of them provide quantifiable risk measures. Recently, a few statistical models, such as the Korean model (6) and pre- and postoperative Early Recurrence After Surgery for Liver Tumor (ERASL) models (7), have been established specifically to predict HCC

Abbreviations

CI = confidence interval, ERASL = Early Recurrence After Surgery for Liver Tumor, HCC = hepatocellular carcinoma, RFS = recurrence-free survival

Summary

Proposed radiomics models integrating pre- and postoperative variables helped predict tumor recurrence before and after resection for early stage hepatocellular carcinoma.

Key Results

- Two radiomics-based models, one integrating preoperative variables and one integrating pre- and postoperative variables, provided better predictive performance (concordance index ≥ 0.77 , $P < .05$ for all), lower prediction error (integrated Brier score ≤ 0.14), and larger net benefits determined with decision curve analysis than rival models without radiomics and widely adopted staging systems.
- The radiomics-based models allow prediction of individual recurrence risk and give three risk strata with low, intermediate, or high risk for recurrence and distinct profiles of recurrent tumor number.

recurrence after liver resection; however, none of them have been validated in patients with early stage HCC, who are the ideal candidates for surgery.

Medical imaging is integral to the routine management of patients with HCC. Conventional imaging evaluation, which refers to the manual assessment of lesions by expert radiologists, relies on semantic features but provides relatively few metrics and discards a wealth of information on tumor heterogeneity (8). In the era of personalized oncology, radiomics has allowed digitally encrypted medical images to be transformed into innumerable quantitative features that provide information on tumor pathophysiology (8–10). The favorable predictive value of radiomics analysis for HCC recurrence has been reported; however, none of these studies have been externally validated and thereafter clinically used yet (11,12).

We aimed to establish recurrence risk models based on radiomics features for HCC meeting the Milan criteria in patients undergoing resection, backed by external validation. We also compared the prognostic and predictive efficacy of these models with that of widely adopted staging systems and other rival recurrence prediction models.

Materials and Methods

Study Participants

The institutional review boards of all participating institutions approved this retrospective multi-institutional study and waived the requirement to obtain written informed consent.

Between February 2009 and December 2016, 377 consecutive patients with HCC meeting the Milan criteria at imaging who underwent curative resection at three institutions were retrospectively recruited (Fig 1). Exclusion criteria were as follows: (a) contrast material-enhanced CT scans of the liver were unavailable, (b) CT was performed more than 1 month before surgery, (c) macrovascular invasion or extrahepatic metastasis was present, (d) patient underwent treatment (ie, repeat liver

resection, local ablation, and transarterial chemoembolization) before surgery, (e) liver resection was performed for ruptured tumor, and (f) clinical-pathologic or follow-up data were not available. A total of 295 patients (median age, 58 years; interquartile range, 50–65 years; 221 men) were enrolled in this study. The development cohort consisted of 177 patients who underwent liver resection at institution 1 (The First Affiliated Hospital of Nanjing Medical University, Nanjing, China). The test cohort consisted of 118 patients who underwent liver resection at two independent institutions (institution 2: Wuxi People's Hospital, Wuxi, China; institution 3: Nanjing First Hospital, Nanjing, China).

Follow-up Surveillance

After resection, patients were followed up and HCC recurrence was screened by means of serum α -fetoprotein level, liver function tests, and contrast-enhanced CT or MRI of the chest and abdomen every 3 months during the first 2 years and then every 6 months thereafter. The data were censored on January 15, 2019. Recurrence-free survival (RFS) was defined as the time from the date of surgery to the date of first recurrence, metastasis, or last follow-up.

CT Technique

Contrast-enhanced CT scans were acquired in the axial plane with 0.75–1.5-mm-thick sections and a 0.75–1.5-mm reconstruction interval. Imaging protocols used at the three institutions are detailed in Appendix E1 and Table E1 (online).

Radiomic Feature Extraction and Selection

The radiomics workflow is shown in Figure 1. All tumors were manually delineated by reader 1 (F.P.Z, with 10 years of experience in liver imaging) on each transverse section from the arterial and portal venous phases by using 3D Slicer (version 4.9.0; <http://www.slicer.org>). The segmented lesions corresponded to the semantically annotated lesions. To capture radiomics features from the tumor periphery, a 4-mm-wide band was generated with automated dilation and shrinkage of tumor boundaries by 2 mm on each side. We explored the stability of each feature extracted from 30 randomly chosen patients; reader 1 repeated tumor segmentation twice in a 1-week period and reader 2 (Q.X., with 20 years of experience in liver imaging) independently performed the segmentation to evaluate test-retest and interreader reproducibility, respectively. The reproducibility was subject to the intraclass correlation coefficient.

Feature extraction and image preprocessing were performed with the open-source Pyradiomics package (version 2.12; <https://pyradiomics.readthedocs.io/en/2.1.2/>). Images were resampled to a voxel size of $1 \times 1 \times 1$ mm to standardize the voxel spacing; voxel intensity values were discretized by using a fixed bin width of 25 HU to reduce image noise and normalize intensities, allowing for a constant intensity resolution across all tumor images (13–15). We extracted 846 radiomic features (19 first-order statistics, 75 texture features, and 752 wavelet decompositions) from each three-dimensional segmentation, giving a total of 3384 features for every lesion (tumor and its periphery in the arterial and portal venous

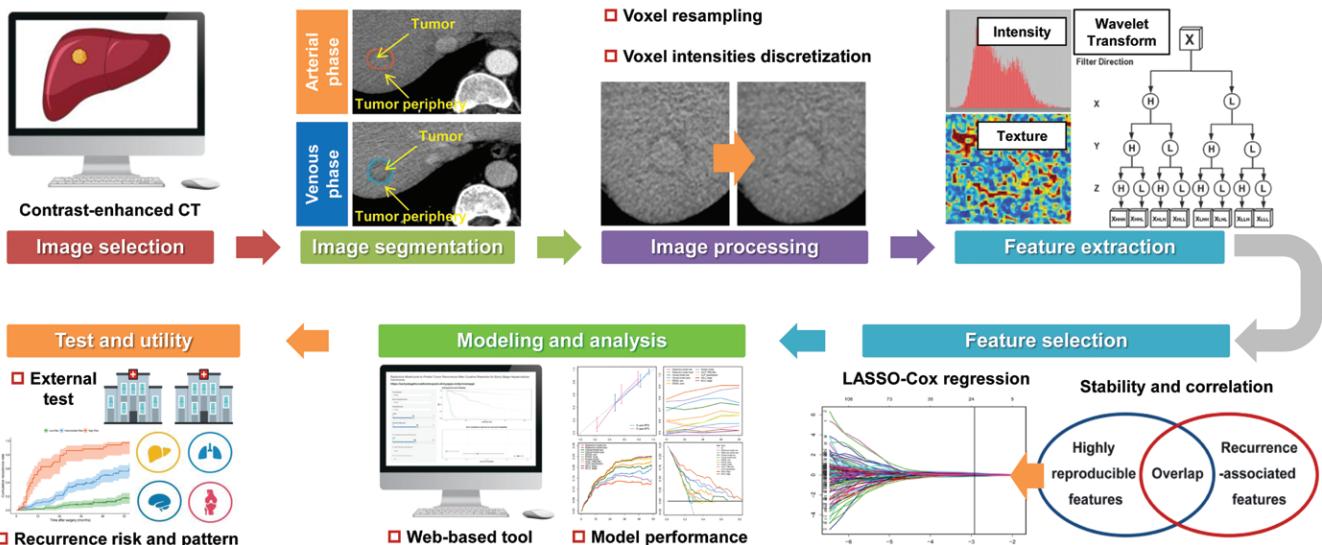
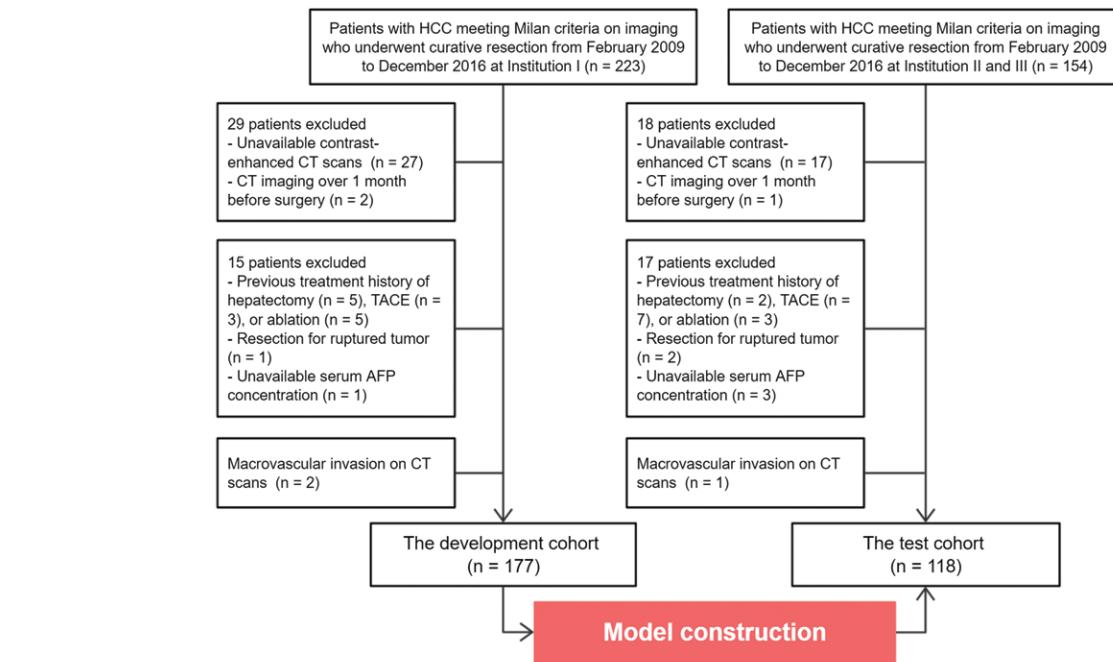


Figure 1: Flowchart (top) shows recruitment pathway for patients. AFP = α -fetoprotein, HCC = hepatocellular carcinoma, TACE = transcatheter arterial chemoembolization. Diagram (bottom) shows radiomics workflow. LASSO = least absolute shrinkage and selection operator.

phases). Feature extraction algorithms are defined in Appendix E1 (online). Values of extracted features for the development cohort were standardized with z scores; feature values of the test cohort were then standardized to z scores by using the mean and standard deviation values derived from the development cohort (Table E2 [online]).

We followed a three-step procedure to identify robust radiomic features. First, features with high stability (intraclass correlation coefficient >0.80) in both test-retest and inter-reader settings were kept for further analysis. Next, univariable Cox regression analysis was applied to select recurrence-related features (with $P < .05$); we then used the least absolute shrinkage and selection operator Cox regression algorithm (16), with penalty parameter tuning conducted by 10-fold cross-validation, to construct a radiomics signature.

CT Evaluation

Two board-certified radiologists (readers 1 and 2) independently reviewed all CT images; they were aware of the diagnosis of HCC but blinded to other information. The radiologists assessed the following features for each patient: (a) tumor number, (b) tumor diameter, (c) liver cirrhosis, (d) tumor margin, (e) capsule appearance, (f) arterial peritumoral enhancement, (g) intratumoral necrosis, and (h) radiogenomic venous invasion, consisting of three separate features (internal arteries, hypoattenuating halo, and tumor-liver difference) that can facilitate preoperative prediction of microvascular invasion according to a three-trait decision tree (17). These semantic features of the largest lesion were recorded for multifocal HCC. Diagnostic criteria and representative CT images are shown in Figure 2 and Appendix E1 (online). The readout sessions were conducted daily

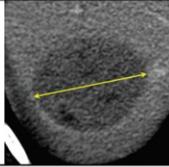
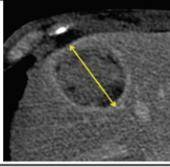
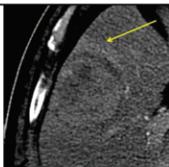
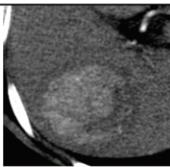
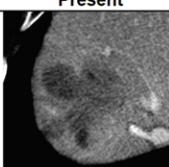
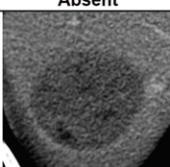
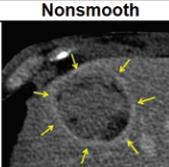
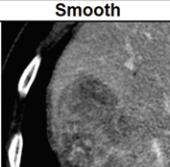
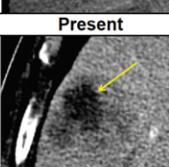
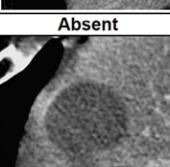
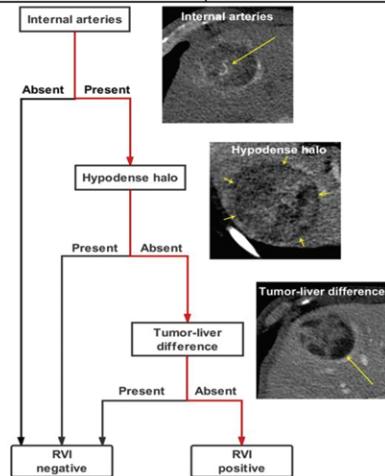
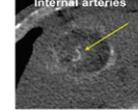
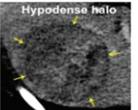
Feature Name	Definition	Example	
Tumor size	The maximum diameter on transverse images		
		Outer edge to outer edge	Include capsule in measurement
Liver cirrhosis	An irregular, nodular or shrunken liver, as well as ascites or evidence of portosystemic collaterals in decompensated stage		
		Present	Absent
Arterial peritumoral enhancement	Detectable portion of enhancement adjacent to the tumor border in the arterial phase, becoming isoattenuating compared with background liver parenchyma in the delayed phase		
		Present	Absent
Tumor margin	Smooth margin, presenting as nodular tumor with smooth border, or nonsmooth margin, presenting as nonnodular tumor with irregular contour		
		Nonsmooth	Smooth
Capsule appearance	Peripheral rim of uniform and smooth hyperenhancement in the portal venous or delayed phase		
		Present	Absent
Intratumoral necrosis	Nonenhancing areas with an attenuation similar to that of gallbladder contents		
		Present	Absent
Radiogenomic venous invasion (RVI)	<p>RVI consists of three imaging features (internal arteries, hypodense halo, and tumor-liver difference) according to a three-trait decision tree.</p> <ul style="list-style-type: none"> Internal arteries: the presence of discrete arteries within the tumor Hypodense halo: a rim of hypoattenuation partially or completely circumscribing the tumor Tumor-liver difference: a focal or circumferential sharp transition in attenuation between the tumor and the adjacent liver parenchyma in the absence of a hypodense halo 		
		  	

Figure 2: Definitions of major semantic features and representative CT images.

Table 1: Baseline Patient Characteristics

Characteristic	Development Cohort (<i>n</i> = 177)	Test Cohort (<i>n</i> = 118)	<i>P</i> Value
Patient demographics			
Age (y)*	58 (49–65)	59 (51–66)	.50
Men	137 (77.4)	84 (71.2)	.26
Liver disease			.70
Hepatitis B virus infection	139 (78.5)	88 (74.6)	
Hepatitis C virus infection	6 (3.4)	4 (3.4)	
Other	32 (18.1)	26 (22.0)	
Laboratory parameters*			
ALT (IU/L)	30.3 (22.0–46.2)	29.4 (22.1–44.5)	.66
Bilirubin (mmol/L)	15.4 (12.2–21.2)	14.6 (11.9–19.4)	.42
Albumin (g/dL)	41.1 (37.1–44.3)	40.2 (36.5–43.2)	.20
Creatinine (μmol/L)	72.4 (63.7–83.7)	70.6 (60.5–84.5)	.61
Prothrombin time (sec)	12.7 (11.8–14.0)	13.3 (11.9–19.9)	.06
Platelet count ($\times 10^9/\text{L}$)	127.0 (84.5–176.5)	129.5 (85.0–189.0)	.60
Child-Pugh grade			.75
A	168 (94.9)	111 (94.1)	
B	9 (5.1)	7 (5.9)	
ALBI grade			.13
1	116 (65.5)	67 (56.8)	
2 or 3	61 (34.5)	51 (43.2)	
MELD score*	4.8 (3.1–7.0)	4.4 (2.9–6.2)	.24
Serum AFP level (ng/mL)*	34.7 (4.6–347.1)	34.4 (3.9–365.8)	.76
CT features			
Tumor diameter (cm)*	3.4 (2.5–4.2)	3.2 (2.5–4.2)	.90
Tumor number			
Solitary	162 (91.5)	104 (88.1)	
Multiple	15 (8.5)	14 (11.9)	
Liver cirrhosis			
Absent	111 (62.7)	74 (62.7)	
Present	66 (37.3)	44 (37.3)	
Capsule appearance			
Incomplete	127 (71.8)	83 (70.3)	
Complete	50 (28.2)	35 (29.7)	
Tumor margin			
Smooth	111 (62.7)	64 (54.2)	
Nonsmooth	66 (37.3)	54 (45.8)	
Arterial peritumoral enhancement			
Absent	165 (93.2)	106 (89.8)	
Present	12 (6.8)	12 (10.2)	
Intratumoral necrosis			
Absent	108 (61.0)	79 (66.9)	
Present	69 (39.0)	39 (33.1)	
Radiogenomic venous invasion			
Negative	145 (81.9)	96 (81.4)	
Positive	32 (18.1)	22 (18.6)	
Radiomics score*	-0.05 (-0.36 to 0.32)	-0.04 (-0.29 to 0.35)	.65
Histologic characteristics			
Edmondson grade			
I-II	122 (68.9)	87 (73.7)	
III-IV	55 (31.1)	31 (26.3)	
Satellite nodules			
Absent	155 (87.6)	106 (89.8)	
Present	22 (12.4)	12 (10.2)	
MVI			.57
Absent	152 (85.9)	104 (88.1)	
Present	25 (14.1)	14 (11.9)	

Table 1 (continues)

Table 1 (continued): Baseline Patient Characteristics

Characteristic	Development Cohort (<i>n</i> = 177)	Test Cohort (<i>n</i> = 118)	<i>P</i> Value
No. of recurrences	84 (47.5)	54 (45.8)	.78
Recurrence-free survival [†]			.33
2-year rate (%)	74.6 (68.4, 81.3)	82.2 (75.6, 89.4)	
5-year rate (%)	51.1 (43.5, 60.1)	56.4 (47.6, 67.0)	
Median time (mo)	62.4 (50.2, NA)	73.7 (56.9, NA)	

Note.—Except where indicated, data are numbers of patients, with percentages in parentheses. AFP = α -fetoprotein, ALBI = albumin-bilirubin, ALT = alanine aminotransferase, MELD = Model for End-Stage Liver Disease, MVI = microvascular invasion, NA = not applicable.

* Data are medians, with interquartile range in parentheses.

[†] Numbers in parentheses are the 95% confidence interval.

Table 2: Multivariable Cox Regression Analysis of Predictors of Recurrence and Creation of Radiomics-based Models in the Development Cohort

Variable	Preoperative Radiomics Model			Postoperative Radiomics Model		
	β	Hazard Ratio	<i>P</i> Value	β	Hazard Ratio	<i>P</i> Value
Natural logarithm of serum AFP level	0.13	1.14 (1.04, 1.26)	.008	0.12	1.13 (1.02, 1.25)	.02
Tumor number (multiple vs solitary)	0.91	2.48 (1.25, 4.93)	.01	0.96	2.62 (1.36, 5.06)	.004
Radiomics signature	1.56	4.75 (3.29, 6.86)	<.001	1.42	4.13 (2.84, 6.01)	<.001
MVI (present vs absent)	NA	NA	NA	0.68	1.97 (1.11, 3.48)	.02
Satellite nodules (present vs absent)	NA	NA	NA	1.13	3.10 (1.82, 5.28)	<.001

Note.—Numbers in parentheses are the 95% confidence interval. Preoperative radiomics model risk score = $0.13 \times \ln(\text{serum AFP level in } \mu\text{g/L}) + 0.91 \times \text{tumor number}$ (0: solitary; 1: multiple) + $1.56 \times \text{radiomics signature}$. Postoperative radiomics model risk score = $0.12 \times \ln(\text{serum AFP level in } \mu\text{g/L}) + 0.96 \times \text{tumor number}$ (0: solitary; 1: multiple) + $1.42 \times \text{radiomics signature} + 0.68 \times \text{MVI}$ (0: absent; 1: present) + $1.13 \times \text{satellite nodules}$ (0: absent; 1: present). The concordance index was 0.77 (standard error of the mean, 0.03) for the preoperative radiomics model and 0.81 (standard error of the mean, 0.02) for the postoperative radiomics model. The Akaike information criterion was 712.57 for the preoperative radiomics model and 696.37 for the postoperative radiomics model. AFP = α -fetoprotein, MVI = microvascular invasion, NA = not applicable.

for 2 weeks. Tumor diameter was recorded as mean value, and any discrepancy in semantic feature assessment was resolved by means of consensus discussion. Interreader variation of semantic features was measured with κ statistics.

Model Development and Validation

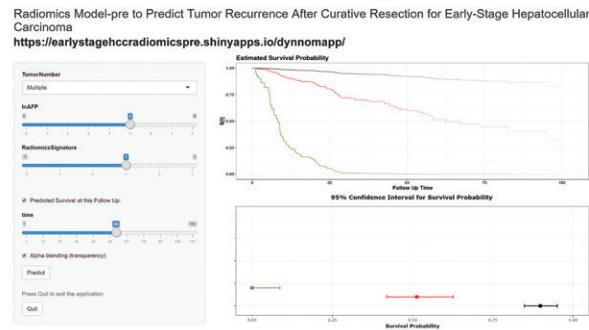
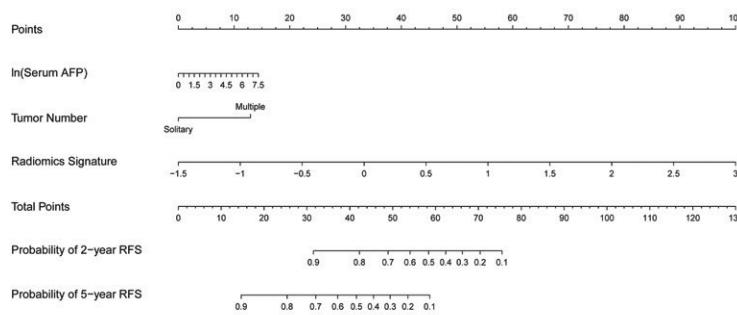
We established two radiomics models for recurrence prediction: The preoperative model included radiomics signature and clinical-radiologic parameters available before surgery; the postoperative model included the aforementioned predictors plus pathologic variables. Correspondingly, two clinical models were generated on the basis of semantic features and parameters available before or after surgery. Predictors of recurrence with statistical significance in the univariable Cox regression analysis were included in the multivariable Cox model; the final model was selected by backward stepwise elimination with Akaike information criteria as the stopping rule (18). The proportional hazards assumption of models was verified by examining the scaled Schoenfeld residual plots. All models were validated in an independent external cohort.

Statistical Analysis

Continuous and categorical variables were compared by using the Mann-Whitney *U* test and χ^2 test, respectively. Natural logarithm transformation of α -fetoprotein values was applied to reduce the effect of small discrepancies. RFS probabilities were estimated by using the Kaplan-Meier method and

compared with the log-rank test. Model discrimination was measured with the concordance index (C index) and compared by using a previously described method (19). Model fit was assessed with a calibration plot by means of 1000 bootstrap resamples. Time-dependent receiver operating characteristic curves and the areas under the curve of each point measured from 12 to 60 months were used to assess prognostic accuracy at different time points (20). The prediction error of models was assessed by using the "Boot632plus" split method with 1000 iterations to calculate estimates of prediction error curves and was summarized as the integrated Brier score, which reflects a weighted average of the squared distances between observed recurrence status and predicted recurrence probability of a model; the integrated Brier score represents a valid measure of overall model performance and can range from 0, for a perfect model, to 0.25, for a noninformative model with a 50% incidence of the outcome (9,21). The clinical utility of the model was evaluated with decision curve analysis (22). Statistical analysis was performed with software (R, version 3.4.4, <http://www.r-project.org>); R packages are listed in Appendix E1 (online). R codes are available at GitHub (https://github.com/radgrady/radiology_Rcode). The optimal cut-point for continuous prognostic markers was obtained by using X-tile software (version 3.6.1; Yale University School of Medicine, New Haven, Conn) (23). $P < .05$ was indicative of a statistically significant difference.

Radiomics Model-pre



Radiomics Model-post

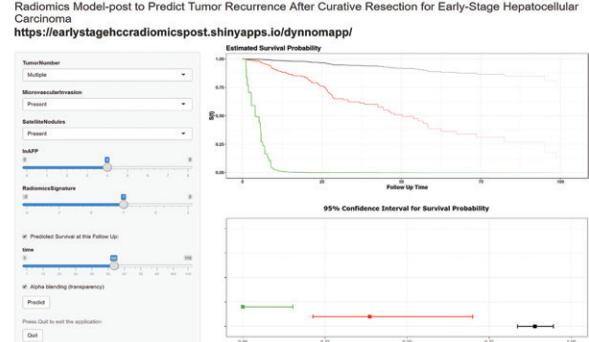
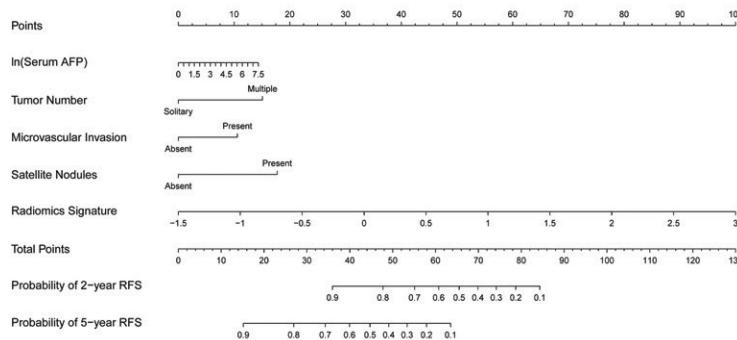


Figure 3: Radiomics-based nomograms and corresponding online calculators to predict recurrence risk before (Model-pre) and after (Model-post) surgery. Online tools are available at <https://earlystagehccradiomicspre.shinyapps.io/dynnomapp/> and <https://earlystagehccradiomicspost.shinyapps.io/dynnomapp/>. AFP = α -fetoprotein, RFS = recurrence-free survival.

Table 3: Prognostic Performance of Radiomics-based Models Compared with Other Models and Staging Systems

Model	Development Cohort				Test Cohort			
	C Index	Time-Dependent AUC	IBS	P Value	C Index	Time-Dependent AUC	IBS	P Value
Preoperative radiomics model	0.77 (0.72, 0.82)	0.82	0.14	Reference	0.78 (0.72, 0.84)	0.84	0.14	Reference
Postoperative radiomics model	0.81 (0.76, 0.85)	0.88	0.12	Reference	0.82 (0.76, 0.87)	0.88	0.13	Reference
Preoperative clinical model	0.68 (0.62, 0.74)	0.72	0.18	<.001*	0.64 (0.57, 0.71)	0.71	0.18	<.001*
Postoperative clinical model	0.73 (0.67, 0.78)	0.76	0.17	<.001†	0.63 (0.55, 0.71)	0.67	0.18	<.001†
Preoperative ERASL model	0.57 (0.51, 0.63)	0.59	0.21	<.001*	0.57 (0.49, 0.64)	0.59	0.20	<.001*
Postoperative ERASL model	0.61 (0.54, 0.68)	0.64	0.19	<.001†	0.59 (0.51, 0.67)	0.63	0.19	<.001†
Korean model	0.53 (0.46, 0.59)	0.56	0.21	<.001†	0.56 (0.48, 0.64)	0.58	0.20	<.001†
BCLC stage	0.54 (0.37, 0.71)	0.52	0.21	.005*	0.54 (0.33, 0.74)	0.51	0.19	.01*
HKLC stage	0.51 (0.34, 0.69)	0.51	0.21	.004*	0.51 (0.34, 0.67)	0.50	0.19	.001*
CLIP classification	0.66 (0.57, 0.76)	0.62	0.20	.01*	0.68 (0.56, 0.80)	0.63	0.19	.04*
AJCC TNM (eighth edition)	0.65 (0.52, 0.79)	0.56	0.21	.009†	0.69 (0.55, 0.84)	0.60	0.20	.03†

Note.—Numbers in parentheses are the 95% confidence interval. The time-dependent AUC represents the median AUC at various time points. All P values were obtained from analyses comparing the C indexes of various models by using the “survcomp” package in R software. AJCC = American Joint Committee on Cancer, AUC = area under the receiver operating characteristic curve, BCLC = Barcelona Clinic Liver Cancer, CLIP = Cancer of the Liver Italian Program, ERASL = Early Recurrence After Surgery for Liver Tumor, HKLC = Hong Kong Liver Cancer, IBS = integrated Brier score.

* P value versus preoperative radiomics model.

† P value versus postoperative radiomics model.

Results

Characteristics of the Study Cohorts

Table 1 shows detailed baseline characteristics of all patients; clinical-radiologic-pathologic characteristics did not differ

between the development and test cohorts. The median duration of follow-up was 50.8 months (interquartile range, 33.5–72.6 months) for the development cohort and 65.4 months (interquartile range, 39.4–78.0 months) for the test cohort. RFS was similar between the two study cohorts ($P = .33$, log-rank test).

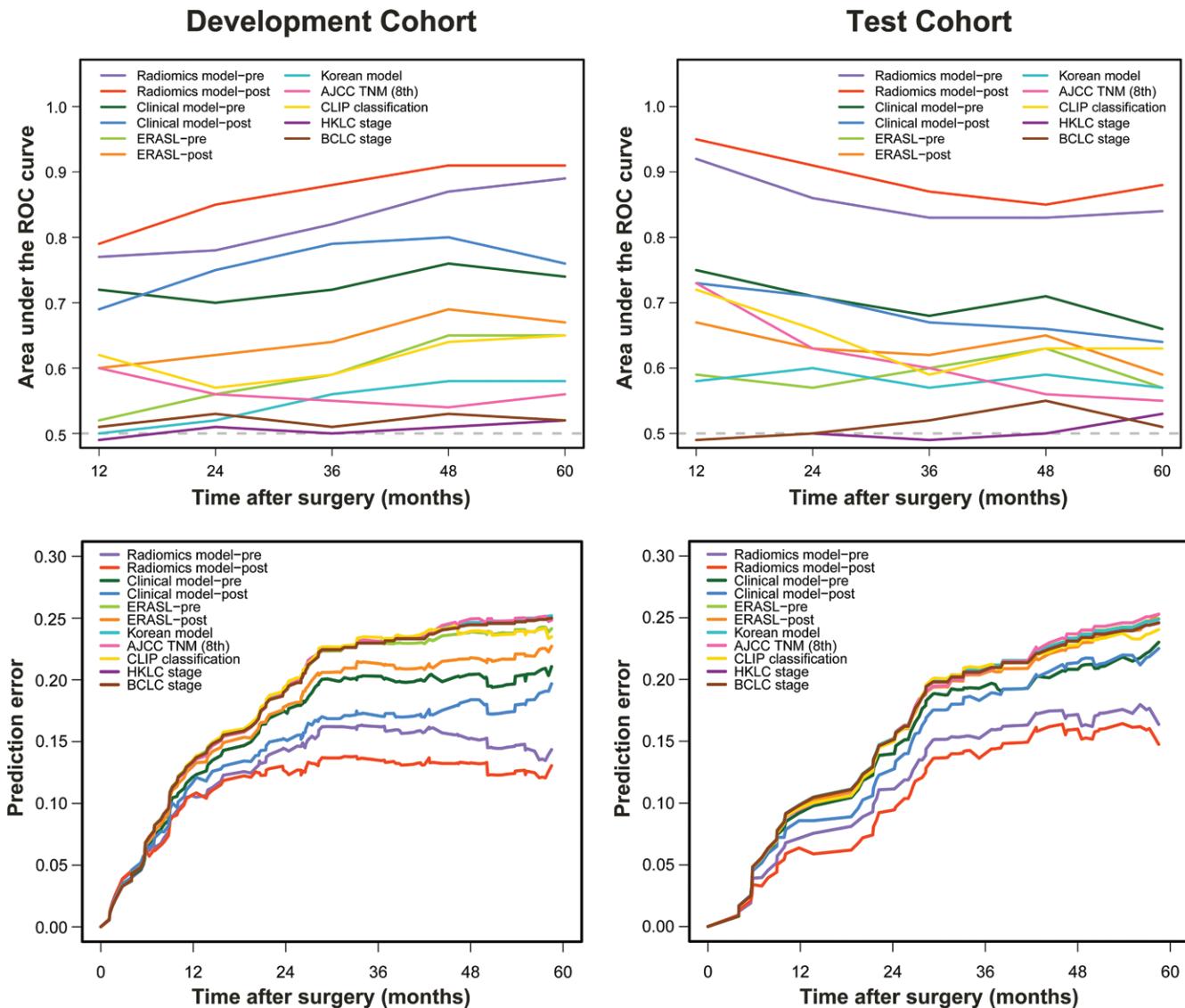


Figure 4: Discriminatory performance and prediction error of all models and systems in development and test cohorts. Graphs show time-dependent areas under the receiver operating characteristic (ROC) curve at various time points (top) and prediction error estimates (bottom) for established models and staging systems. AJCC = American Joint Committee on Cancer, BCLC = Barcelona Clinic Liver Cancer, CLIP = Cancer of the Liver Italian Program, ERASL = Early Recurrence After Surgery for Liver Tumor, HKLC = Hong Kong Liver Cancer.

Radiomic Feature Selection and Signature Construction

Among 2865 radiomic features with high stability, 273 features that helped predict recurrence in univariable Cox regression were identified. The least absolute shrinkage and selection operator Cox regression model was then used to select 20 features to derive a radiomics signature (Fig E1 [online]). The formula for radiomics signature and the distribution of corresponding score are presented in Appendix E1 (online) and Figure E2 (online). The radiomics signature indicated favorable prediction of HCC recurrence with a C index of 0.73 (95% confidence interval [CI]: 0.67, 0.78) in the development cohort and 0.74 (95% CI: 0.68, 0.80) in the test cohort.

Prediction Models Development and Validation

In the development cohort, 10 variables were predictive of HCC recurrence at univariable analysis (Fig E3 [online]). With use of stepwise multivariable analysis with the lowest

Akaike information criteria score, independent predictors were identified for the preoperative and postoperative radiomics models (Table 2). Two radiomics models that integrated corresponding independent predictors were developed and presented as two nomograms as well as web-based tools to provide individualized risk estimates (Fig 3). There were no significant violations of the proportional hazards assumption, assessed by scaled Schoenfeld residuals against time for each predictor in the radiomics models (Fig E4 [online]). Similarly, clinical pre- and postoperative models were produced according to the formulas shown in Table E3 (online). Agreement between the two radiologists for the two semantic features in clinical models was excellent ($\kappa = 0.819$ for tumor margin and 0.926 for intratumoral necrosis).

The C indexes of the preoperative radiomics model for predicting HCC recurrence in the development and test cohorts were 0.77 (95% CI: 0.72, 0.82) and 0.78 (95% CI: 0.72, 0.84),

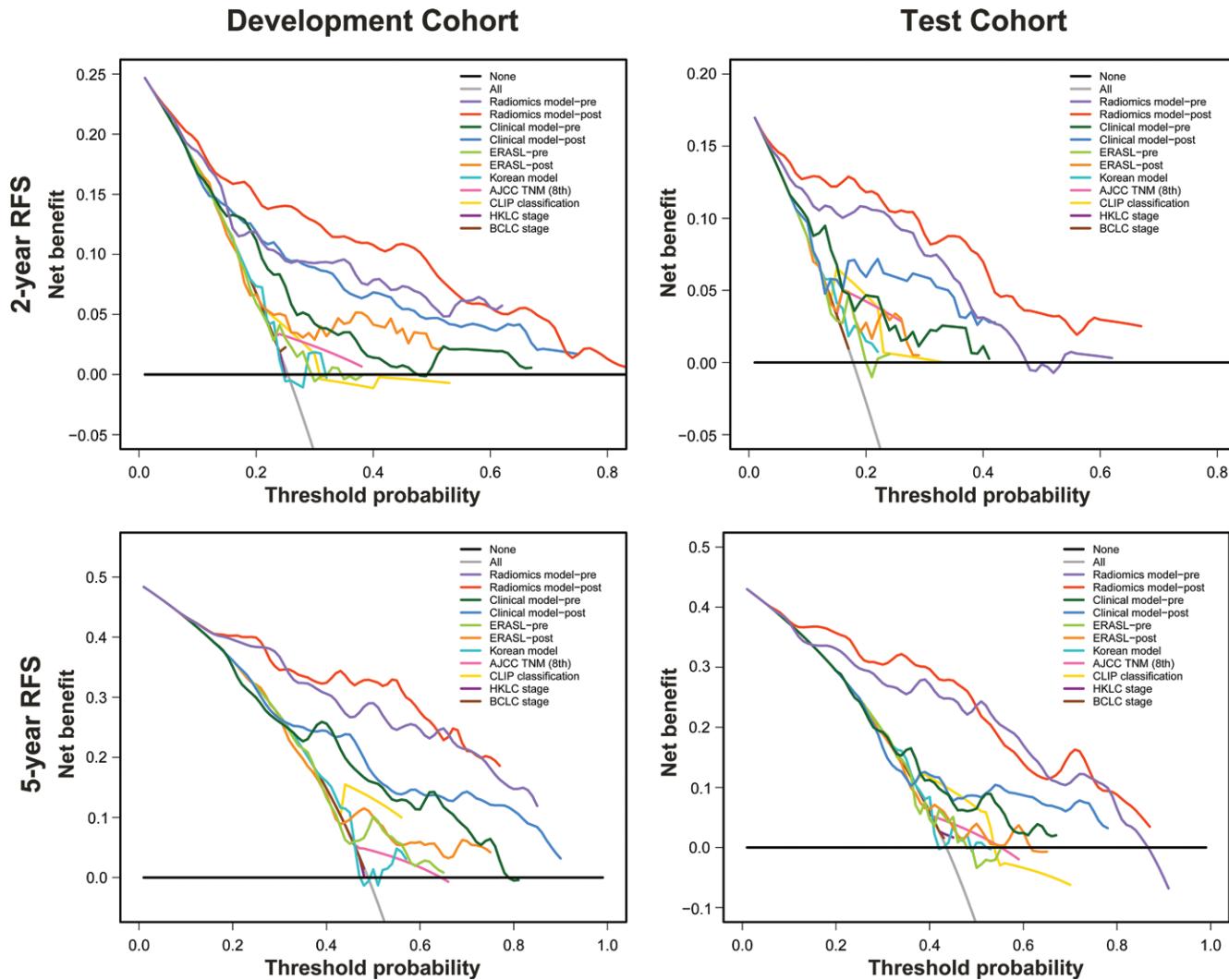


Figure 5: Decision curves for recurrence-free survival (RFS) at 2 years (top) and 5 years (bottom) obtained by using established models and staging systems in the development and test cohorts. The y-axis measures the net benefit, which was calculated by summing the benefits (true-positive results) and subtracting the harms (false-positive results), weighting the latter by a factor related to the relative harm of an undetected tumor compared with the harm of unnecessary treatment. Both radiomics-based models provided the highest net benefit compared with rival models, widely adopted staging systems, and simple strategies (eg, follow-up of all patients or no patients across the majority of the range of reasonable threshold probabilities at which a patient would choose to undergo imaging follow-up). AJCC = American Joint Committee on Cancer, BCCLC = Barcelona Clinic Liver Cancer, CLIP = Cancer of the Liver Italian Program, ERASL = Early Recurrence After Surgery for Liver Tumor, HKLC = Hong Kong Liver Cancer.

respectively. The prognostic performance of the preoperative radiomics model was superior ($P < .05$) to that of preoperative clinical model, preoperative ERASL model, and other staging systems (Barcelona Clinic Liver Cancer, Hong Kong Liver Cancer, Cancer of the Liver Italian Program) in both study cohorts (Table 3). The postoperative radiomics model yielded the best discriminatory ability, with C index values of 0.81 (95% CI: 0.76, 0.85) in the development cohort and 0.82 (95% CI: 0.76, 0.87) in the test cohort; its prognostic accuracy was higher ($P < .05$) than that of the postoperative clinical model, postoperative ERASL model, Korean model, and TNM system. Calibration plots for two radiomics-based models in predicting 2- and 5-year RFS demonstrated that the model-predicted RFS was well calibrated with the Kaplan-Meier–observed RFS in both cohorts (Fig E5 [online]).

Using time-dependent receiver operating characteristic curve analysis, we found that two radiomics-based models

improved prediction of HCC recurrence compared with rival models and staging systems at various time points in both study cohorts (Fig 4). Complete details of time-dependent areas under the receiver operating characteristic curves for each model or system are reported in Table E4 (online). Prediction error curves of all models and staging systems are presented in Figure 4. The integrated Brier scores for pre- and postoperative radiomics models were 0.14 and 0.12, respectively, in the development cohort and 0.14 and 0.13 in the test cohort; both radiomics-based models provided more precise prognostication of RFS than did rival models and systems in both study cohorts (Table 3). Decision curve analysis graphically demonstrated that two radiomics-based models provided larger net benefit across the range of reasonable threshold probabilities compared with other models, staging systems, and simple

Table 4: Median RFS and Cumulative Tumor Recurrence Rates according to Each Risk Group Defined by the Two Radiomics-based Models

Model and Group	No. of Patients	Median RFS (mo)	TRR at 2 Years (%)	TRR at 5 Years (%)	Hazard Ratio	P Value
Preoperative radiomics model						
Development cohort						
Low-risk group	68	98.7 (95.1, -NA)	7.4 (0.9, 13.4)	17.9 (6.7, 27.8)	1	
Intermediate-risk group	86	46.0 (33.2, 73.7)	26.7 (16.8, 35.5)	61.2 (46.3, 71.9)	3.87 (2.31, 6.49)*	<.001*
High-risk group	23	10.1 (7.0, 24.2)	73.9 (48.1, 86.9)	100.0 (NA, NA)	14.08 (6.34, 31.26)†	<.001†
Test cohort						
Low-risk group	44	NA (92.4, NA)	2.3 (0.0, 6.6)	17.3 (4.7, 28.2)	1	
Intermediate-risk group	54	62.4 (46.0, NA)	13.0 (3.5, 21.5)	48.8 (30.5, 62.3)	3.03 (1.56, 5.89)*	<.001*
High-risk group	20	11.0 (7.0, 27.8)	65.0 (36.4, 80.7)	93.3 (58.7, 98.9)	9.86 (4.16, 23.35)†	<.001†
Postoperative radiomics model						
Development cohort						
Low-risk group	104	NA (95.1, NA)	7.7 (2.4, 12.7)	23.1 (13.2, 31.9)	1	
Intermediate-risk group	41	36.9 (28.3, 56.9)	29.3 (13.9, 41.9)	83.1 (51.6, 94.1)	11.30 (5.47, 23.37)*	<.001*
High-risk group	32	12.1 (8.9, 21.4)	78.1 (57.9, 88.6)	100.0 (NA, NA)	5.50 (2.96, 10.24)†	<.001†
Test cohort						
Low-risk group	69	NA (92.4, NA)	1.4 (0.0, 4.2)	16.8 (6.6, 25.9)	1	
Intermediate-risk group	28	38.3 (28.2, NA)	14.3 (0.3, 26.3)	71.9 (45.0, 85.7)	11.27 (4.86, 26.16)*	<.001*
High-risk group	21	10.1 (7.8, 24.2)	76.2 (48.8, 88.9)	100.0 (NA, NA)	4.46 (2.08, 9.59)†	<.001†

Note.—Numbers in parentheses are the 95% confidence interval. NA = not applicable, RFS = recurrence-free survival, TRR = tumor recurrence rate.

* In comparison to the low-risk group.

† In comparison to the intermediate-risk group.

strategies (ie, follow-up of all patients or no patients) in both study cohorts (Fig 5).

Recurrence Risk Stratification

With use of 0.2 and 1.7 as cutoff scores derived from X-tile analysis (Fig E6 [online]) of the development cohort (which correspond to total points of 37 and 57 in the nomogram, respectively), the preoperative radiomics model identified three risk categories of recurrence (median RFS, 98.7 months, 46.0 months, and 10.1 months for low-, intermediate-, and high-risk patients, respectively; $P < .001$) (Table 4). Cumulative 2- and 5-year recurrence rates were 7.4% and 17.9%, respectively, for low-risk patients, 26.7% and 61.2% for intermediate-risk patients, and 73.9% and 100.0% for high-risk patients (Fig 6). Similar results were obtained for the postoperative radiomics model with use of 0.9 and 1.8 as cutoff values (which correspond to total points of 47 and 62 in the nomogram, respectively) by using X-tile analysis (Fig E6 [online]). In the test cohort, three distinct prognostic strata were confirmed with the two radiomics-based models ($P < .001$ for all) (Table 4, Fig 6).

During follow-up, there were 84 (47.5%, 84 of 177) and 54 (45.8%, 54 of 118) documented recurrences in the development and test cohorts, respectively. The number of recurrent HCCs and corresponding treatments differed among the three risk categories predicted by either the pre- or postoperative radiomics models in both study cohorts ($P < .05$ for all). In brief, multiple recurrent nodules were more commonly detected in intermediate- and high-risk patients than in low-risk patients; a higher proportion of low-risk patients received

potentially curative therapy (liver transplant, repeat liver resection, or ablation) for recurrent HCC compared with intermediate- and high-risk patients (Tables E5, E6 [online]).

Discussion

The aim of our study was to develop and validate recurrence risk models based on contrast-enhanced CT radiomics for early stage hepatocellular carcinoma (HCC). We concluded that radiomic features incorporated into a recurrence prediction model exhibited improved performance in early stage HCCs compared with other models and systems. Specifically, in this multi-institutional study, the two radiomics-based models provided better prognostic ability (C index ≥ 0.77 ; $P < .05$ for all), lower prediction error (integrated Brier score ≤ 0.14), and better clinical usefulness compared with rival models and staging systems; the two models were well calibrated. In addition, both models could successfully categorize early stage HCC into three recurrence risk subgroups with distinct profiles of recurrent tumor number. Similar performance of radiomics-based models in the development cohort and the test cohort from other institutions suggested the reproducibility and reliability of the proposed models.

Resection for early stage HCCs is still plagued by high recurrence rate, and about half of the patients in our study developed tumor relapse. Early stage HCC is heterogeneous, with diverse outcomes. Gene signatures may refine prognosis scoring; however, they are not used in routine clinical care (1). Conversely, the power and potential of medical image data are increasingly recognized in the field of oncology (8). In our study, we established four image-based prediction models for

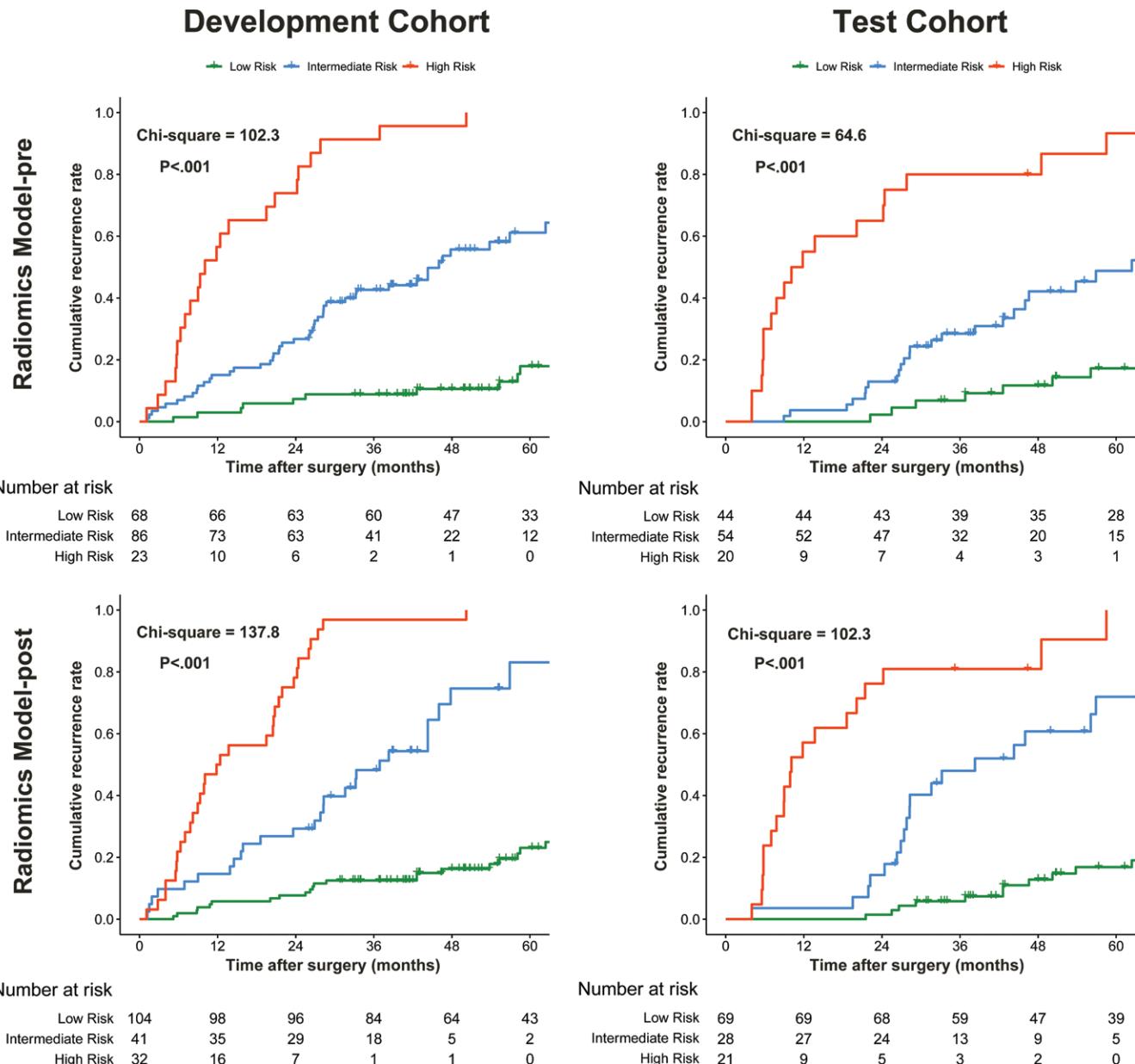


Figure 6: Graphs show cumulative rates of tumor recurrence according to three risk strata defined by the radiomics model before (Model-pre) and after (Model-post) surgery in the development and test cohorts.

HCC recurrence. Two semantics-based clinical models showed relatively higher discrimination ability than ERASL models, the Korean model, and widely adopted staging systems, suggesting the great value of radiologic features to complement clinical-pathologic features for recurrence prediction. Alternatively, radiomic features enable quantification of tumor heterogeneity by representing the spatial arrangement of imaging voxels with signal intensity variations (8–10). The two radiomics-based models used in our study significantly outperformed the two semantics-based clinical models in discriminatory ability. Given that engineered radiomics is highly dependent on image acquisition parameters, we applied voxel intensity discretization and voxel size resampling to reduce the dependency of differences in image specifications, in accordance with recent CT radiomics studies (13–15), and therefore

to improve the translational potential of radiomics-based models. Wavelet-based features achieved the highest weights in our signature, in accordance with previous reports (15,24). These features may further reflect the spatial heterogeneity of a tumor and its periphery at multiple scales. Accordingly, radiomics features can provide more detailed information on tumor biology, as well as tumor microenvironment, that are complementary to visual features.

No accepted adjuvant therapies have been demonstrated to reduce recurrence in resected HCCs; however, patients at high risk for recurrence are potential candidates for clinical trials of adjuvant therapy (1–3). Conversely, liver transplant is the most effective treatment to prevent recurrence, whereas primary liver resection for initially resectable HCC followed by salvage liver transplant in cases of

transplantable tumor recurrence is a highly applicable strategy. It allows the optimal allocation of scarce organs and provides survival outcomes similar to those of upfront liver transplant, but the low actual transplantability rate remains the Achilles heel of this strategy (25,26). An alternative approach is to enlist patients at high risk for recurrence for liver transplant before recurrence detection (27). Nevertheless, challenge remains in the individualized prediction of recurrence risk. Our radiomics models could affect both the use of adjuvant treatment and the liver transplant strategy by individualizing management according to the three risk profiles for recurrence. By using the preoperative model, we advocate upfront liver transplant for high-risk patients; the postoperative model allows prophylactic and pre-emptive enlistment of high-risk patients for salvage liver transplant before the recurrence is identified. In addition, the two radiomics models may facilitate individualized surveillance policy. Specifically, low-risk patients may receive no adjuvant therapy and a less intensive surveillance regimen, even within the first 2 years after surgery, given their 2-year cumulative recurrence rate of less than 8%. Intermediate-risk patients should undergo intensive surveillance lasting for 5 years because of the high risk for recurrence even after 2 years, accompanied by adjuvant systemic therapies.

Our study has limitations. First, it was a retrospective analysis that suffers from inherent biases, although external validation was conducted to improve the reliability. Second, our study was conducted in China, and most patients had hepatitis B virus-related HCC. Third, all scans were obtained with single-vendor CT scanners (Siemens) at all three institutions; our results require further validation with other CT vendors to check for the generalizability. Fourth, computing z scores for the test cohort by using statistics from the development cohort may have resulted in a slightly optimistic assessment of model performance. Finally, the association of radiomic features with genomic patterns was not investigated.

In conclusion, radiomic features offer notable advantages over existing prognostic sources for early stage hepatocellular carcinoma (HCC). Our radiomics models with pre- and postresection features are powerful and externally validated tools to predict tumor recurrence, thus providing an unprecedented opportunity to improve clinical decision support for patients with early stage HCC. Further studies are required to explore the generalized utility of our models and translate our results into clinical practice.

Author contributions: Guarantors of integrity of entire study, all authors; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, M.Y.W., X.C.L.; clinical studies, F.P.Z., Q.X.; statistical analysis, Q.X., K.W., M.Y.W., W.W.T.; and manuscript editing, G.W.J., X.C.L., X.H.W.

Disclosures of Conflicts of Interest: G.W.J. disclosed no relevant relationships. F.P.Z. disclosed no relevant relationships. Q.X. disclosed no relevant relationships. K.W. disclosed no relevant relationships. M.Y.W. disclosed no relevant relationships.

W.W.T. disclosed no relevant relationships. X.C.L. disclosed no relevant relationships. X.H.W. disclosed no relevant relationships.

References

1. Villanueva A. Hepatocellular carcinoma. *N Engl J Med* 2019;380(15):1450–1462.
2. Heimbach JK, Kulik LM, Finn RS, et al. AASLD guidelines for the treatment of hepatocellular carcinoma. *Hepatology* 2018;67(1):358–380.
3. European Association for the Study of the Liver. EASL Clinical Practice Guidelines: Management of hepatocellular carcinoma. *J Hepatol* 2018;69(1):182–236 [Published correction appears in *J Hepatol* 2019;70(4):817].
4. 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(12):1622–1629.
5. Wang JH, Wang CC, Hung CH, Chen CL, Lu SN. Survival comparison between surgical resection and radiofrequency ablation for patients in BCLC very early/early stage hepatocellular carcinoma. *J Hepatol* 2012;56(2):412–418 <https://doi.org/10.1016/j.jhep.2011.05.020>.
6. Shim JH, Jun MJ, Han S, et al. Prognostic nomograms for prediction of recurrence and survival after curative liver resection for hepatocellular carcinoma. *Ann Surg* 2015;261(5):939–946.
7. Chan AWH, Zhong J, Berhane S, et al. Development of pre and post-operative models to predict early recurrence of hepatocellular carcinoma after surgical resection. *J Hepatol* 2018;69(6):1284–1293.
8. Bi WL, Hosny A, Schabath MB, et al. Artificial intelligence in cancer imaging: Clinical challenges and applications. *CA Cancer J Clin* 2019;69(2):127–157.
9. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 2017;14(12):749–762.
10. Ji GW, Zhang YD, Zhang H, et al. Biliary tract cancer at CT: a radiomics-based model to predict lymph node metastasis and survival outcomes. *Radiology* 2019;290(1):90–98.
11. Kim S, Shin J, Kim DY, Choi GH, Kim MJ, Choi JY. 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(13):3847–3855.
12. Zheng BH, Liu LZ, Zhang ZZ, et al. Radiomics score: a potential prognostic imaging feature for postoperative survival of solitary HCC patients. *BMC Cancer* 2018;18(1):1148.
13. Sun R, Limkin Ej, Vakalopoulou M, et al. A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study. *Lancet Oncol* 2018;19(9):1180–1191.
14. Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 2014;5(1):4006.
15. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res* 2017;77(21):e104–e107.
16. Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med* 1997;16(4):385–395.
17. Banerjee S, Wang DS, Kim HJ, et al. A computed tomography radiogenomic biomarker predicts microvascular invasion and clinical outcomes in hepatocellular carcinoma. *Hepatology* 2015;62(3):792–800.
18. Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Contr* 1974;19(6):716–723.
19. Schröder MS, Culhane AC, Quackenbush J, Haibe-Kains B. survcomp: an R/Bioconductor package for performance assessment and comparison of survival models. *Bioinformatics* 2011;27(22):3206–3208.
20. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* 2000;56(2):337–344.
21. Cohen ME, Ko CY, Bilmoria KY, et al. Optimizing ACS NSQIP modeling for evaluation of surgical quality and risk: patient risk adjustment, procedure mix adjustment, shrinkage adjustment, and surgical focus. *J Am Coll Surg* 2013;217(2):336–46.e1.
22. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006;26(6):565–574.
23. Camp RL, Dolled-Filhart M, Rimm DL. X-tile: a new bio-informatics tool for biomarker assessment and outcome-based cut-point optimization. *Clin Cancer Res* 2004;10(21):7252–7259.
24. Liang W, Yang P, Huang R, et al. A combined nomogram model to preoperatively predict histologic grade in pancreatic neuroendocrine tumors. *Clin Cancer Res* 2019;25(2):584–594.
25. Bhangui P, Allard MA, Vibert E, et al. Salvage versus primary liver transplantation for early hepatocellular carcinoma: do both strategies yield similar outcomes? *Ann Surg* 2016;264(1):155–163.
26. de Haas RJ, Lim C, Bhangui P, et al. Curative salvage liver transplantation in patients with cirrhosis and hepatocellular carcinoma: an intention-to-treat analysis. *Hepatology* 2018;67(1):204–215.
27. Ferrer-Fábregas J, Forner A, Lluch A, et al. Prospective validation of ab initio liver transplantation in hepatocellular carcinoma upon detection of risk factors for recurrence after resection. *Hepatology* 2016;63(3):839–849.