

This article is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND) (<http://www.karger.com/Services/OpenAccessLicense>). Usage and distribution for commercial purposes as well as any distribution of modified material requires written permission.

Original Paper

Deep Learning Radiomics Based on Contrast-Enhanced Ultrasound Might Optimize Curative Treatments for Very-Early or Early-Stage Hepatocellular Carcinoma Patients

Fei Liu^{a–c} Dan Liu^a Kun Wang^{b, c} Xiaohua Xie^a Liya Su^a
Ming Kuang^{a, d} Guangliang Huang^a Baogang Peng^d Yuqi Wang^{b, c}
Manxia Lin^a Jie Tian^{b, e} Xiaoyan Xie^a

^aDepartment of Medical Ultrasonics, Institute of Diagnostic and Interventional Ultrasound, The First Affiliated Hospital of Sun Yat-sen University, Guangzhou, China; ^bCAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Beijing, China; ^cDepartment of the Artificial Intelligence Technology, University of Chinese Academy of Sciences, Beijing, China; ^dDepartment of Liver Surgery, The First Affiliated Hospital of Sun Yat-sen University, Guangzhou, China; ^eBeijing Advanced Innovation Center for Big Data-Based Precision Medicine, Beihang University, Beijing, China

Keywords

Contrast-enhanced ultrasound · Hepatocellular carcinoma · Radiomics · Radiofrequency ablation · Surgical resection

Abstract

Background: We aimed to evaluate the performance of a deep learning (DL)-based Radiomics strategy designed for analyzing contrast-enhanced ultrasound (CEUS) to not only predict the progression-free survival (PFS) of radiofrequency ablation (RFA) and surgical resection (SR) but also optimize the treatment selection between them for patients with very-early or early-stage hepatocellular carcinoma (HCC). **Methods:** We retrospectively enrolled 419 patients examined by CEUS within 1 week before receiving RFA or SR (RFA: 214, SR: 205) from January 2008 to 2016. Two Radiomics signatures were constructed by the Radiomics model R-RFA and R-SR to stratify PFS of different treatment groups. Then, RFA and SR nomograms were built by incorporating Radiomics signatures and significant clinical variables to achieve individualized 2-year PFS prediction. Finally, we applied both Radiomics models and both nomograms to each enrolled patient to investigate whether there were space for treatment optimization and how much prognostic improvement could be expected. **Results:** R-RFA and R-SR showed

F.L., D.L., and K.W. contributed equally to this work and should be considered as co-first authors.

Manxia Lin, MD, PhD and Xiaoyan Xie, MD, PhD, Department of Medical Ultrasonics, Institute of Diagnostic and Interventional Ultrasound, The First Affiliated Hospital of Sun Yat-sen University 58 Zhongshan Second Road, Guangzhou, Guangdong 510080 (China) linmxia@mail.sysu.edu.cn and xxy1992sys@163.com

Jie Tian, PhD
CAS Key Laboratory of Molecular Imaging
Institute of Automation, Chinese Academy of Sciences
95 Zhongguancun East Road, Beijing 100191 (China)
tian@iee.org

remarkable discrimination (C-index: 0.726 for RFA, 0.741 for SR). RFA and SR nomograms provided good 2-year PFS prediction accuracy and good calibrations. We identified 17.3% RFA patients and 27.3% SR patients should swap their treatment, so their average probability of 2-year PFS would increase 12 and 15%, respectively. **Conclusions:** The proposed Radiomics models and nomograms achieved accurate preoperative prediction of PFS for RFA and SR, and they could facilitate the optimized treatment selection between them for patients with very-early or early-stage HCC.

© 2020 The Author(s)
Published by S. Karger AG, Basel

Introduction

Surgical resection (SR) and radiofrequency ablation (RFA) are the 2 main curative strategies for early-stage hepatocellular carcinoma (HCC) [1]. Extensive randomized clinical trials have been conducted to compare the long-term survivals of RFA and SR in the treatment of early-stage HCC, reporting significantly contrasting conclusions [2–4]. Whether SR or RFA is the superior treatment strategy for individuals remains a controversial issue, since RFA was proposed as a therapy alternative for small HCC by the American Association for the Study of Liver Diseases in 2005 [5]. Previous studies have reported that the hepatic insufficiency, tumor burden, and location are significant prognostic factors associate with the choice of RFA and SR [6–8]. However, the utility of these clinical factors is still limited and ambiguous in daily clinical practice. Therefore, new individualized predictive methods are urgently needed to identify the optimal candidates who will benefit the most from RFA or SR, preoperatively.

Dynamic contrast-enhanced ultrasound (CEUS) is a promising work-up that can offer outstanding temporal resolution in tracing the microcirculation perfusion of HCC [9]. The inflow blood system, tumor sinusoids, and outflow blood system change drastically during multistep hepatocarcinogenesis, and CEUS has great advantages in visualizing the difference of hemodynamic perfusions in different HCCs [10]. Previous studies revealed that certain quantifiable patterns of CEUS were related to treatment outcomes of RFA or SR [10–12]. Therefore, we hypothesize that the heterogeneity of HCC dynamic behavior is associated with the tumor response to different treatments. If such heterogeneity can be interpreted quantitatively from CEUS, we may obtain extra precision of prognostic prediction regarding each HCC patient, so that individualized treatment selection between RFA and SR can be achieved and implemented into daily practice.

To better interpret CEUS, we developed a deep learning (DL)-based Radiomics strategy. DL-based Radiomics is an emerging technology aimed to assist clinical decision making by learning prognosis related features from medical images, which may not be perceived by human eyes or human-defined features [13]. To the best of our knowledge, no previous study has investigated whether a Radiomics approach would achieve effective treatment selection between RFA and SR for an individual patient with very-early or early-stage HCC.

Materials and Methods

Study Design and Patients

The protocol of this single-center retrospective study was approved by the Institutional Review Board and requirement for informed consent was waived. From January 2008 to 2016, 470 consecutive patients with HCC who underwent CEUS examination within 1 week before RFA or SR were recruited (RFA: $n = 243$, SR: $n = 227$). The diagnosis of HCC was with reference to the diagnostic criteria issued by the European Association for the Study of the Liver [1]. The inclusion criteria were as follows: (1) solitary primary tumor with a maximum diameter ≤ 5.0 cm; (2) good liver function with Child-Pugh class A; (3) performance status

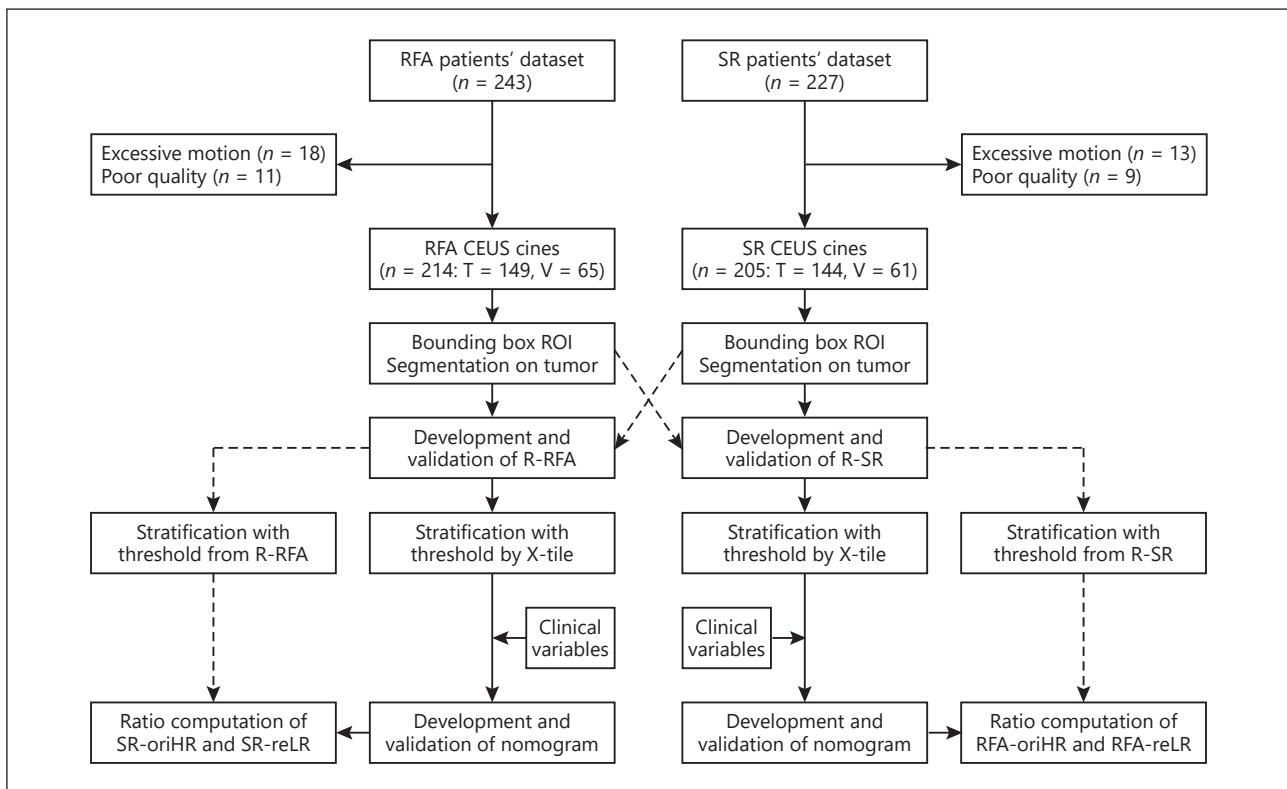


Fig. 1. Patient enrollments and DL-based Radiomics model application flowchart. R-RFA and R-SR models were built using CEUS cines from the RFA group and the SR group, respectively. RFA and SR nomograms were built by incorporating clinical variables and Radiomics signatures extracted from R-RFA and R-SR, respectively. Cross stratifications were conducted by applying R-RFA and R-SR into CEUS cines from swapped patient groups, respectively. Then, the individualized PFS prediction was compared between the original and restratified patient subgroups for each of the RFA and SR groups. RFA, radiofrequency ablation; SR, surgical resection; CEUS, contrast-enhanced ultrasound; T, training cohort; V, validation cohort; ROI, region of interest.

Eastern Cooperative Oncology Group score 0 or 1; (4) more than 24 months follow-up time for patients without tumor progression. The exclusion criteria were as follows: (1) poor quality of CEUS cines (e.g., entire tumor and surrounding liver parenchymal did not clearly display on ultrasound image at the same time); (2) excessive motion during CEUS examination. Demographic information, imaging examination, and pretreatment clinical baseline characteristics were collected from the Institutional Picture Archiving and Communication System (PACS[®]; Carestream Health, Toronto, Canada). The final decision whether to undergo RFA or SR was made by a multidisciplinary team of specialists from different departments based on tumor size, tumor location, liver function, performance status, and patient's will. Finally, the CEUS cines of 419 patients (RFA: $n = 214$, SR: $n = 205$) acquired pretreatment were retrospectively analyzed (Fig. 1).

CEUS Data Acquisition

CEUS was part of routine work-up for patients with HCCs in our department. Examinations were performed by 1 of the 2 radiologists with over 10 years of experience in liver CEUS. Two types of ultrasound instruments, Philips iU22 and Toshiba Aplio, were used. The scanning acoustic window and depth were adjusted to simultaneously display the longitudinal section of the tumor and its surrounding liver parenchyma on 1 scanning section. Gain, dynamic range, mechanical index, output power, and focal zone were optimized to ensure sufficient tissue cancellation with maintenance of adequate penetration. A volume of 2.4 mL of the second-generation contrast agent (SonoVue[®]; Bracco Imaging, Milan, Italy) was injected within 1 s via the elbow, followed by a 5-mL saline flush. Timer was activated promptly from the beginning of

SonoVue injection. The transducer was maintained in a stable position for 90 s to examine the tumor and the surrounding liver parenchyma, and every 10 s thereafter until clearance of contrast agent. Contrast clips were stored in Dicom format. The mechanical index was 0.08 for Philips iU22 and was 0.05–0.08 for Toshiba Aplio. The frame rate was 20–25 frames per second in both types of equipments. We resampled and unified the frame rates of all CEUS cines into 2 frames per second in the course of preprocessing to eliminate the influence of different frame rates regarding to the performance of models, as well as reducing the risk of overfitting for DL.

RFA and SR

All RFA procedures were performed percutaneously by 1 of the 2 radiologists with >10 years of experience in tumor ablation. Conscious analgesia-sedation plus local anesthesia was used during ablation. The Cool-tip electrodes (maximum power: 200 W, 17-gauge, internally cooling cycle, active tip: 2–3 cm; Valleylab®, USA) were inserted into the target tumor under ultrasound guidance. For each patient, 1–3 electrodes were used. Ablation was performed using preset power and duration parameters that were based on the recommended protocol. The electrodes were tracked back or repositioned in an attempt to ablate the whole tumor with a safe margin of ≥ 5 mm. The insertion tract was ablated before withdrawal. In cases in which an insufficient margin was detected on CEUS examination performed immediately after RFA, an additional RFA session was attempted on the same or the following day. SR was carried out under general anesthesia using a right subcostal incision with a midline extension. In this study, anatomic partial hepatectomy was routinely performed in most patients (192/205, 93.7%) with a resection margin of at least 10 mm. However, nonanatomical resection was performed in a small part of patients (13/205, 6.3%) who might not access to a sufficient surgical margin because of advanced liver cirrhosis or a peripherally located tumor. Pringle's maneuver was routinely used with a clamp/unclamp time of 10/5 min. Hemostasis on the surface of the liver was done with a bipolar electric coagulator, argon beam coagulator, titanium clips, tie or suturing, and some hemostats.

Follow-Up Protocol

Patients were followed-up at 1, 3, 6, 9, and 12 months, and every 3–6 months thereafter. Serum alpha-fetoprotein and imaging examination (contrast-enhanced computed tomography or contrast-enhanced magnetic resonance imaging) was performed. Technique efficacy of ablation was assessed at 1 month after the treatment. Successful ablation was defined as tumor was completely ablated with at least 5 mm ablation margin. The primary endpoint was progression-free survival (PFS). PFS was calculated from the date RFA or SR was performed to the radiological identification of tumor progression (local tumor progression, new intrahepatic tumor, vascular invasion, or distant organ metastases – whichever came first). PFS was censored at the date of death from any causes or the date of the last follow-up visit for progression-free patients. Local tumor progression was defined as the tumor was adjacent to the ablation/resection margin within 1.0 cm.

Tumor Annotation and Preprocessing

CEUS cines were annotated manually by an ultrasound doctor with 6 years of experience analyzing liver CEUS cines. The doctor was blinded to the PFS observation and clinical characteristics during the process of tumor annotation. ITK-SNAP was used as annotation software [14] (<http://www.itksnap.org>). We further modified the software by automatically creating a bounding box based on the doctor's manual annotations. Doctors only needed to give a quick and rough annotation of the tumor. Then, the topmost, bottommost, leftmost, and rightmost boundary points of tumor were automatically defined to construct the bounding box. As long as the major tumor area was reasonably contained inside the box, the annotation was considered acceptable. After that, the region of interest (ROI) was defined by expanding the bounding box 1 cm outward, so that the entire tumor area and some surrounding liver parenchyma were contained in it. We adopted 3-min CEUS for analysis.

In this study, 2 PFS-prediction models were established (R-RFA for the RFA group and R-SR for the SR group; Fig. 1). The image annotation method for R-RFA and R-SR was the same. A bounding box was manually drawn on a CEUS frame with distinct tumor margin to include the whole tumor area and part of surrounding liver parenchyma (Fig. 2a, red box). Then, the bounding box was automatically generated on every frame of the CEUS cine. Manual correction of the box location might be applied on certain frames, if the doctor considered it was necessary.

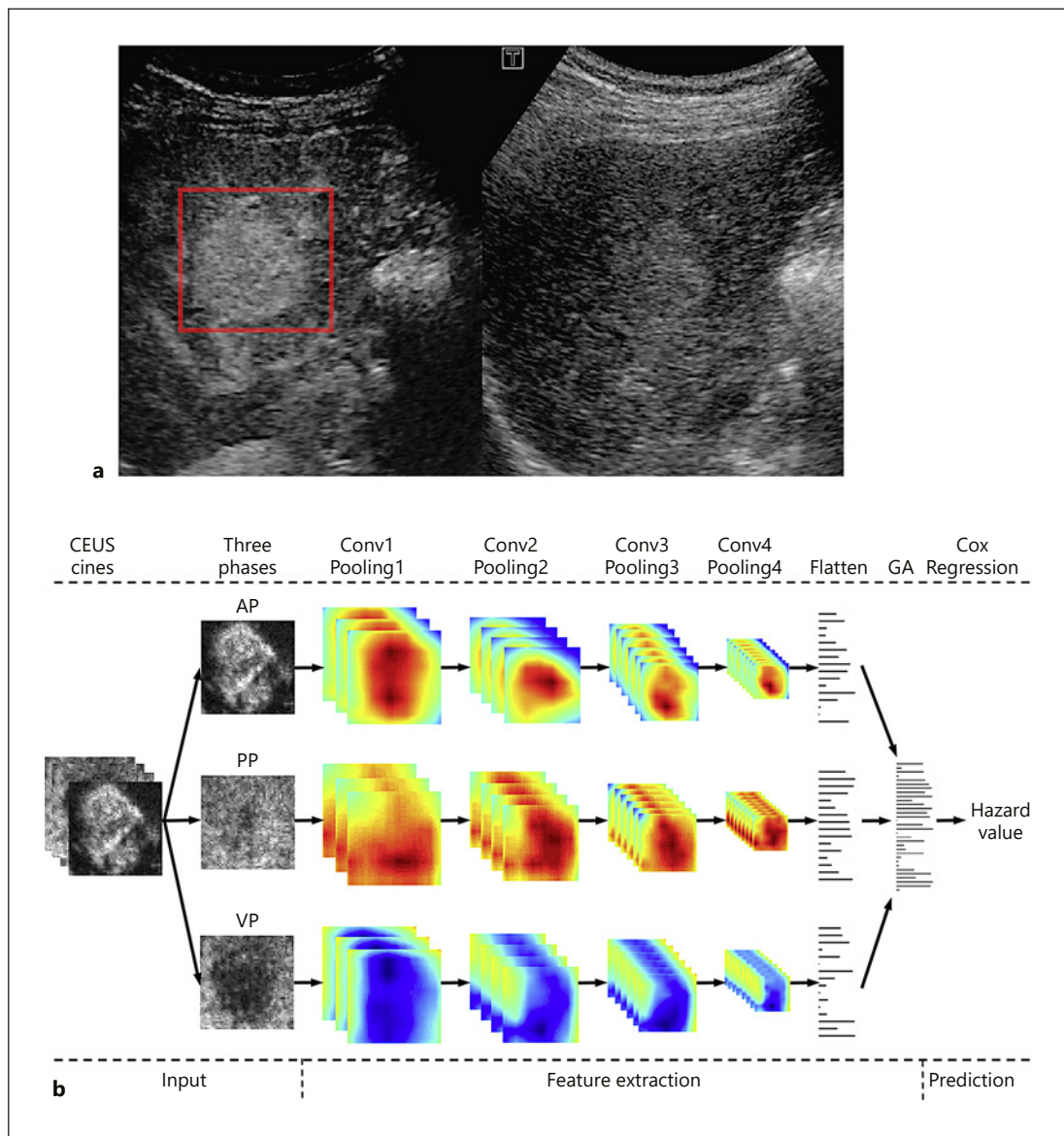


Fig. 2. Illustration of ROI annotation in CEUS cines and the design of R-RFA and R-SR models. **a** An example of the red bounding box ROI drawn in one CEUS frame. Note that the frame was converted from color to grayscale to be loadable for ITK-SNAP. **b** The designed DL network structure of R-RFA and R-SR. Input step: a three-dimensional ROI (2D in space and 1D in time) of CEUS cines was fed into the CNN model to obtain the discriminative features by automatic feature extraction. Feature extraction step: the features extracted from AP, PP, and VP were then aggregated into one global feature set that described the characteristics of the entire CEUS cine. Prediction step: finally, survival hazard was obtained by the Cox-regression algorithm as the output, which essentially was the estimated risk of a patient receiving RFA or SR. CEUS, contrast-enhanced ultrasound; AP, arterial phase; PP, portal phase; VP, venous phase; Flatten, a DL network layer to transform a 3D feature map into a 1D feature; GA, global aggregation.

Development and Validation of DL-Based Radiomics Models

Before the Radiomics model building, CEUS cines of enrolled patients in each of RFA and SR groups were randomly divided into training and validation cohorts by a ratio of 2:1 (Fig. 1). The random division process was repeated until there was no significant difference in clinical characteristics between training and vali-

dation cohorts (Table 2). Subsequently, in both RFA and SR, CEUS cines and follow-up data of training cohorts were used to train and optimize Radiomics models (R-RFA and R-SR) to predict the PFS hazard and probability. Then, CEUS cines and follow-up data of validation cohorts were used to validate and quantify prediction performances of these 2 Radiomics models.

R-RFA and R-SR were developed using a Convolutional Neural Network (CNN)-based Cox proportional hazards regression algorithm (Cox-CNN) [15, 16]. This artificial intelligence (AI) algorithm was designed to automatically extract thousands of spatio-temporal CEUS features from the annotated ROI in arterial, portal, and venous phases, and then intelligently learn key features specific for predicting PFS in RFA or SR (Fig. 2b). We also visualized the features of DL-based Radiomics to further explore how DL-based Radiomics models interpret CEUS cines for prognostic prediction. Please find detailed description of Cox-CNN and feature visualization in the supplementary methods (see www.karger.com/doi/10.1159/000505694 for all online suppl. material).

For each patient, the survival hazard was computed as the output of R-RFA or R-SR, and it was denoted as the Radiomics signature. For patients in RFA or SR, the stratification threshold based on Radiomics signatures was evaluated in the corresponding training cohort by X-tile [17], then both training and validation cohorts were stratified into low- and high-risk subgroups. Different PFS behaviors in stratified subgroups were plotted in Kaplan-Meier survival curves [18]. Harrell's concordance index (C-index) was used to measure the performance of R-RFA and R-SR [19]. A C-index score around or above 0.70 signifies satisfactory performance [13].

Development and Validation of Nomograms for Individualized Prediction

To build easy-to-use and comprehensive nomograms for individualized prediction in RFA and SR, we assessed the impacts of clinical variables and Radiomics signatures (calculated from R-RFA and R-SR) for prognostic prediction through multivariable Cox regression analysis [20]. Candidates of clinical variables included age, cirrhosis, etiology, alpha-fetoprotein, alanine aminotransferase (ALT), white blood cell, platelet count (PLT), prothrombin time, total bilirubin, albumin (ALB), perivascular location, located in periphery of liver (subcapsular location or adjacent to diaphragm, gastrointestinal tract, and gallbladder), tumor size, and albumin-bilirubin scores [21, 22]. The selection was performed using Akaike's information criterion in a backward stepwise approach [23]. The clinical variables and Radiomics signature with significant prognostic values ($p < 0.05$) were selected to build the nomogram for individualized prediction of 2-year PFS in RFA or SR [23, 24].

Calibration curves of RFA and SR nomograms in their corresponding training and validation cohorts were plotted to compare their predictions and real observations in 2-year PFS, respectively [24]. Closer distance between a calibration curve and the diagonal indicates a higher accuracy. The Hosmer-Lemeshow test was applied to further assess the quality of calibration, and an insignificant test statistic implied perfect calibration [25]. Harrell's C-index was also measured. In addition, decision curve analysis was applied to compare the net clinical benefits between using and not using Radiomics signatures at different thresholds [26].

Cross Stratification Using R-RFA and R-SR in Swapped Groups

To optimize the treatment selection between RFA and SR for all enrolled HCC patients, after the original low- and high-risk stratification in the RFA group (RFA-oriLR and RFA-oriHR) using R-RFA, this group was re-stratified into low- and high-risk subgroups (RFA-reLR and RFA-reHR) using R-SR. Similarly, the SR group was also stratified into low- and high-risk subgroups using R-SR (SR-oriLR and SR-oriHR) and R-RFA (SR-reLR and SR-reHR), separately. The Mann-Whitney U test was applied to compare the difference of hazards, so that the effectiveness of such re-stratifications using swapped Radiomics models was investigated.

After that, in either RFA or SR group, overlapped and nonoverlapped patients between the original and second stratifications using different Radiomics models were categorized in 4 subgroups. Then, we focused on the nonoverlapped patients between RFA-oriHR and RFA-reLR in the RFA group, as well as between SR-oriHR and SR-reLR in the SR group. Both RFA and SR nomograms were applied to calculate the 2-year PFS probabilities of each individual patient in these 2 nonoverlapped subgroups, so that their prognostic predictions corresponding to RFA and SR were quantitatively compared using *t* test.

Statistical Analysis

Descriptive statistics were summarized as mean with SD or mean with 95% CI. R-RFA and R-SR model building and evaluation were conducted using Python (version 2.7, <https://www.python.org/>), PyTorch

Table 1. Preoperative clinical characteristic of patients in the RFA and SR groups

Characteristic	RFA (n = 214), n (%)	SR (n = 205), n (%)	p value
Age, years, mean ± SD (95% CI)	56.2±11.0 (29–83)	53.9±11.2 (2–78)	0.302
≤60	138 (64.5)	142 (69.3)	
>60	76 (35.5)	63 (30.7)	
Gender			0.250
Male	190 (88.8)	174 (84.9)	
Female	24 (11.2)	31 (15.1)	
Etiology			0.971
HBV	181 (84.6)	180 (87.8)	
HCV	10 (4.7)	7 (3.4)	
Fatty liver	2 (0.9)	4 (2.0)	
Unknown/other	21 (9.8)	14 (6.8)	
Liver cirrhosis			0.118
Yes	121 (56.5)	100 (48.8)	
No	93 (43.5)	105 (51.2)	
Performance status			0.630
0	168 (78.5)	165 (80.5)	
1	46 (21.5)	40 (19.5)	
AFP, ng/mL			0.881
<20	96 (44.9)	81 (39.5)	
20–200	70 (32.7)	43 (21.0)	
≥200	48 (22.4)	81 (39.5)	
ALT, U/L	41.7±33.5 (4.3–224.0)	46.3±52.4 (11.0–448.0)	0.339
TBIL, μmol/L	17.9±10.1 (4.0–72.2)	16.43±8.13 (3.9–77.8)	0.102
PT, seconds	13.1±1.5 (10.9–23.2)	12.83±1.68 (10.7–31.0)	0.084
ALB, g/L			0.290
<35	39 (18.2)	29 (14.1)	
≥35	175 (81.8)	176 (85.9)	
WBC, ×10 ⁹ /L, mean ± SD (95% CI)	5.7±3.3 (1.1–46.0)	5.9±2.8 (1.1–29.4)	0.505
PLT, ×10 ⁹ /L, mean ± SD (95% CI)	154.4±76.2 (17.0–468.0)	165.9±64.3 (19.0–482.0)	0.100
Tumor size, cm			0.141
≤2	113 (52.8)	91 (44.4)	
2–5	101 (47.2)	110 (53.6)	
Tumor location			0.967
Right lobe	180 (84.1)	150 (73.2)	
Left lobe	20 (9.3)	50 (24.4)	
Bilobar	14 (6.6)	5 (2.4)	
Perivascular location			0.975
Yes	44 (20.6)	43 (21.0)	
No	170 (79.4)	162 (79.0)	
Located in periphery of liver			0.106
Yes	55 (25.7)	65 (31.7)	
No	159 (74.3)	140 (68.3)	
ALBI scores			0.952
1	112 (52.34)	102 (49.76)	
2	98 (45.79)	100 (48.78)	
3	4 (1.87)	3 (1.46)	

RFA, radiofrequency ablation; SR, surgical resection; HBV, hepatitis B virus; HCV, Hepatitis C virus; AFP, alpha-fetoprotein; ALT, alanine aminotransferase; TBIL, total bilirubin; PT, prothrombin time; ALB, albumin; WBC, white blood cell; PLT, platelet count; Perivascular location, Yes if tumor is adjacent to ≥3 mm vessel and No if not; Located in periphery of liver, Yes if tumor is within 5 mm close to the liver capsule, gallbladder, and/or gastrointestinal tract, and No if not; ALBI scores, albumin-bilirubin scores.

(version 0.4.0, <https://pytorch.org/>) and lifeline (version 0.19.5, <https://lifelines.readthedocs.io/>). Tools related with Radiomics nomograms establishment and evaluation were R software (version 3.4.4, <https://www.r-project.org/>) with packages survival, rms, ResourceSelection, and compareC. Student *t* test or Mann-Whitney test, as appropriate, and χ^2 test were used to compare continuous and categorical variables, respectively, which were performed using GraphPad Prism[®] 5.0 (GraphPad Software, La Jolla, CA, USA). All statistical tests were two-sided. Differences were considered significant at $p < 0.05$.

Results

Clinical Characteristics

There were 419 patients (RFA: 214, SR: 205) enrolled in this study. All patients achieved successful ablation in the RFA group. All clinical variables had no significant difference between the RFA and SR group ($p > 0.05$; Table 1). The median follow-up periods for the RFA and SR groups were 92.1 and 66.9 months, respectively. In the RFA group, 92 patients (92/214, 43.0%) suffered progression, including local tumor progression (12, 5.6%), intrahepatic metastasis (64, 29.9%), vascular invasion (4, 1.9%), and extrahepatic metastasis (12, 5.6%). In the SR group, 90 patients (90/205, 43.9%) suffered progression, including local tumor progression (0, 0%), intrahepatic metastasis (75, 36.6%), vascular invasion (6, 2.9%), and extrahepatic metastasis (9, 4.4%). There was no significant difference in PFS between RFA and SR (median PFS: RFA group, 81.6 months; SR group, 59.7 months, $p = 0.12$). Patients were randomly divided into training and validation cohorts in both groups, which were 149/65 for the RFA group and 144/61 for the SR group. Meanwhile, we ensured that there were no significant differences in PFS and clinical variables between the training and validation cohorts, neither for the RFA nor the SR group (both $p > 0.05$; Table 2).

Development and Validation of DL-Based Radiomics Models

We built the DL-based R-RFA model and R-SR model to estimate the hazards of PFS, and then stratified patients into low- and high-risk subgroups (online suppl. Table S1). The optimum stratification thresholds generated by the X-tile were -0.1 for R-RFA and 0.7 for R-SR. For RFA, the ratios of high-risk subgroups in the training and validation cohorts were 23.49 and 35.38%, respectively. For SR, the ratios were 48.61 and 37.10%, respectively. After the stratification in RFA and SR groups, Kaplan-Meier curves showed significant difference between low- and high-risk subgroups in the training and validation cohorts (Fig. 3a–d, $p < 0.005$ in all the cohorts). The C-indexes of R-RFA in the training and validation cohorts were 0.754 (95% CI 0.701–0.808) and 0.726 (95% CI 0.650–0.802), accompanied with hazard ratios of 5.543 (95% CI 3.381–7.086) and 5.384 (95% CI 3.018–7.821), respectively. Likewise, the R-SR model achieved a C-index of 0.787 (95% CI 0.735–0.838) for the training cohort and 0.741 (95% CI 0.640–0.852) for the validation cohort, with hazard ratios of 2.903 (95% CI 2.223–3.792) and 3.477 (95% CI 2.512–4.250).

Visualization of DL-Based Radiomics Models

To better perceive and understand the functional mechanism of R-RFA and R-SR, we visualized their internal Radiomics feature maps in real time and found certain human-recognizable feature patterns from these DL algorithms. In the RFA group, the R-RFA model tended to pay great attention to the microbubble wash-in of the entire tumor area during the arterial and portal phases (online suppl. Fig. S2a, red areas), but such attention vanished during the venous phase (online suppl. Fig. S2a, blue areas). Therefore, R-RFA achieved prognostic prediction by looking into all pixels inside the tumor during AP and PP. In contrast, the R-SR model turned its eyes to the tumor periphery instead of the entire tumor area, where red

Table 2. Baseline characteristics in the training and validation cohorts

Characteristic	RFA (n = 214), n (%)		SR (n = 205), n (%)		p
	T (n = 149)	V (n = 65)	T (n = 144)	V (n = 61)	
Age, years, mean ± SD (95% CI)					
≤60	56.36±10.72 (29–83)	55.87±11.62 (32–80)	54.04±11.11 (20–78)	53.61±11.32 (33–74)	0.801
>60	96 (64.42)	42 (64.62)	102 (70.83)	40 (65.57)	
Gender	53 (35.58)	23 (35.38)	42 (29.17)	21 (34.43)	0.675
Male	134 (89.93)	56 (86.15)	121 (84.03)	53 (86.89)	
Female	15 (10.07)	9 (13.85)	23 (15.97)	8 (13.11)	0.446
Liver cirrhosis					
Yes	83 (55.70)	38 (58.46)	73 (50.69)	27 (44.26)	
No	66 (44.30)	27 (41.54)	71 (49.31)	34 (55.74)	0.336
Performance status					
0	115 (77.2)	53 (81.5)	113 (78.5)	52 (85.2)	
1	34 (22.8)	12 (18.5)	31 (21.5)	9 (14.8)	0.106
AFP, ng/mL					
<20	70 (46.98)	26 (40.00)	65 (45.14)	15 (26.20)	
20–200	46 (30.87)	24 (36.92)	38 (26.39)	6 (9.84)	0.362
≥200	33 (22.15)	15 (23.08)	41 (28.47)	40 (65.56)	0.163
ALT, U/L	40.27±31.02 (4.3–224)	45.13±38.27 (11–194)	48.21±49.95 (11–448)	55.54±58.05 (14–411)	0.697
TBLi, μmol/L	17.56±10.03 (4–72.2)	18.82±10.49 (5.2–62.9)	16.93±8.74 (6.6–77.8)	15.25±6.29 (3.9–38.0)	0.899
PT, s	13.07±1.45 (10.9–19.5)	13.17±1.67 (10.9–23.2)	12.79±1.12 (10.7–17.9)	12.89±2.56 (10.7–31.0)	0.200
ALB, g/L					
<35	31 (20.82)	8 (12.30)	21 (14.48)	8 (13.11)	0.418
≥35	118 (79.18)	57 (87.70)	123 (85.52)	53 (86.89)	0.139
WBC, ×10 ⁹ /L, mean ± SD (95% CI)	5.54±2.01 (1.19–12.07)	6.31±5.33 (2.38–46.0)	6.09±3.17 (1.18–29.48)	5.54±1.62 (2.09–10.37)	
PLT, ×10 ⁹ /L, mean ± SD (95% CI)	156.95±80.47 (17–468)	148.65±65.07 (61–324)	168.29±66.86 (38–482)	160.56±57.34 (19–318)	0.457
Tumor size, cm					
≤2	111 (74.49)	44 (67.69)	40 (27.78)	17 (27.86)	
2–5	38 (25.51)	21 (31.31)	104 (72.22)	44 (72.14)	0.708
Tumor location					
Right lobe	125 (83.90)	55 (84.61)	109 (75.69)	41 (67.21)	
Left lobe	13 (8.72)	7 (10.77)	33 (22.92)	17 (27.86)	0.414
Bilobar	11 (7.38)	3 (4.62)	2 (1.39)	3 (4.93)	
Perivascular location					
Yes	28 (18.79)	16 (24.62)	29 (20.14)	14 (22.95)	
No	121 (81.21)	49 (75.38)	115 (79.86)	47 (77.05)	0.333
Located in periphery of liver					
Yes	37 (24.83)	18 (27.69)	43 (29.86)	22 (36.07)	
No	112 (75.17)	47 (72.31)	101 (70.14)	39 (63.93)	
ALBI scores					
1	79 (53.03)	33 (50.77)	68 (47.22)	34 (55.74)	
2	67 (44.97)	31 (47.69)	74 (51.39)	26 (42.63)	
3	3 (2.00)	1 (1.54)	2 (1.39)	1 (1.63)	

RFA, radiofrequency ablation; SR, surgical resection; AFP, alpha-fetoprotein; ALT, alanine aminotransferase; TBLi, total bilirubin; PT, prothrombin time; ALB, albumin; WBC, white blood cell; PLT, platelet count; Perivascular location, Yes if tumor is adjacent to ≥3 mm vessel and No if not; Located in periphery of liver, Yes if tumor is within 5 mm close to the liver capsule, gallbladder, and/or gastrointestinal tract, and No if not; ALBI scores, albumin-bilirubin scores.

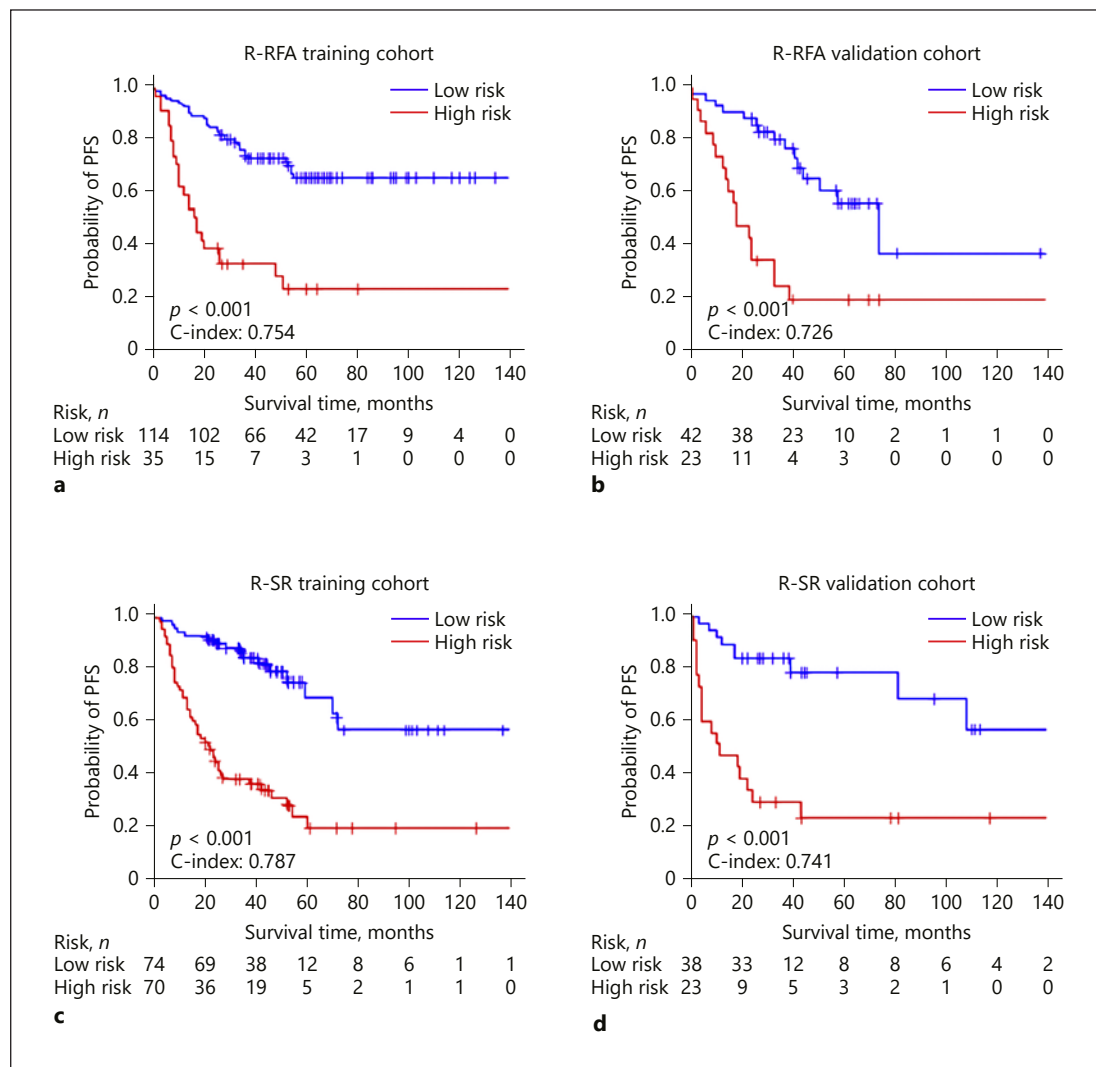
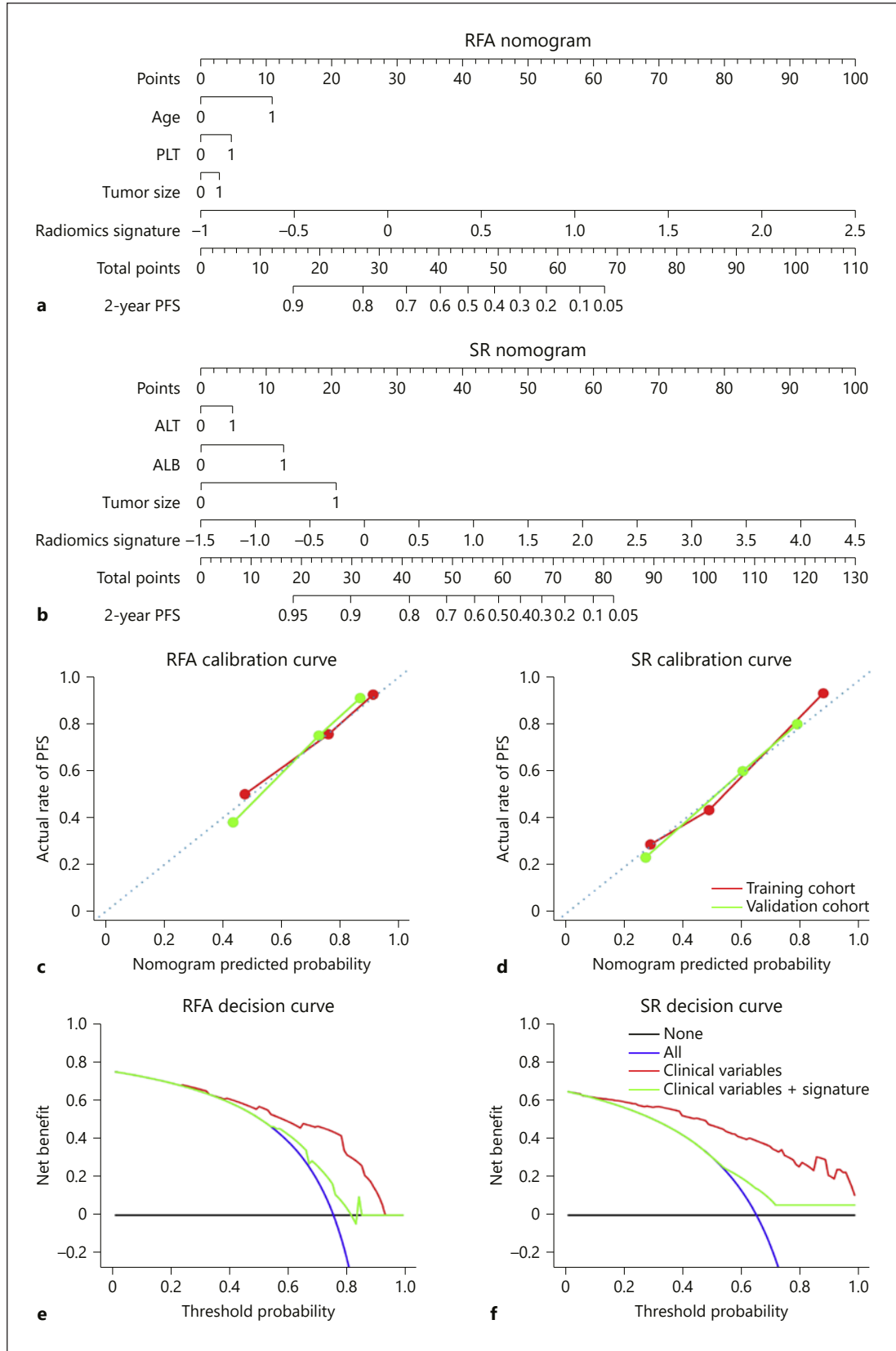


Fig. 3. Kaplan-Meier survival curves and time-dependent ROC curves. Kaplan-Meier survival curves of R-RFA in the training and validation cohorts (**a, b**) and for R-SR (**c, d**). The short vertical lines indicate censored data. RFA, radiofrequency ablation; PFS, progression-free survival; SR, surgical resection.

Fig. 4. Constructed nomograms to estimate individualized PFS for RFA and SR, along with calibration and decision curves. **a, b** Nomograms for predicting the probability of a patient having 2-year PFS after receiving RFA or SR. **a** Age = 0 means age ≤ 60 years, age = 1 means age > 60 years, PLT = 0 means $100 \times 10^9 \leq \text{PLT} \leq 300 \times 10^9 \text{ L}$, PLT = 1 means $\text{PLT} < 100 \times 10^9 \text{ L}$, tumor size = 0 means tumor size $\leq 2 \text{ cm}$, tumor size = 1 means tumor size $> 2 \text{ cm}$. **b** ALT = 0 means $\text{ALT} \leq 40 \text{ U/L}$, ALT = 1 means $\text{ALT} > 40 \text{ U/L}$, ALB = 0 means $\text{ALB} \geq 35 \text{ g/L}$, ALB = 1 means $\text{ALB} < 35 \text{ g/L}$, tumor size = 0 means tumor size $\leq 2 \text{ cm}$, tumor size = 1 means tumor size $> 2 \text{ cm}$. **c, d** Calibration curves given by the RFA and SR nomograms by using their corresponding training and validation cohorts. **e, f** Decision curve analysis for RFA and SR nomograms whether integrating their Radiomics signatures (red curves) or not (green curves). In both plots, black lines represent the assumption that no patients showed PFS, and blue curves represent the assumption that all patients showed PFS. The y-axis measures the net benefit. The threshold probability in x-axis was where the expected benefit of treatment is equal to the expected benefit of avoiding treatment. RFA, radiofrequency ablation; PFS, progression-free survival; SR, surgical resection; PLT, platelet count; ALT, alanine aminotransferase; ALB, albumin.

(For figure see next page.)



pixels aggregated into a doughnut-like shape in each frame, and this pattern was consistent in all 3 phases (online suppl. Fig. S2b).

Development and Validation of Nomograms for Individualized Prediction

Multivariable Cox regression analysis identified age ($p = 0.013$), PLT ($p = 0.026$), tumor size ($p = 0.045$), and Radiomics signatures ($p < 0.0001$) as independent predictors for RFA prognosis. However, independent variables selected for SR prognosis were ALT ($p = 0.039$), ALB ($p = 0.013$), tumor size ($p = 0.043$), and Radiomics signatures ($p < 0.0001$; online suppl. Table S2). Based on these variables, individualized PFS prediction models for RFA and SR were developed and presented as nomograms (Fig. 4a, b).

The calibration curves of both nomograms exhibit good agreement between the prediction results and the observations in the training and validation cohorts (Fig. 4c, d). The Hosmer-Lemeshow test also indicated that there was no significant deviation from the perfect fit ($p = 0.330$ and 0.479 for RFA, $p = 0.209$ and 0.403 for SR, in training and validation cohorts, respectively).

We further applied RFA and SR nomograms for patient stratifications in both RFA and SR groups regarding to their 2-year PFS predictions. The optimum stratification thresholds were 1.6 for RFA and 1.2 for SR. Their corresponding Kaplan-Meier curves are presented in online supplementary Figure S3a–d. All stratifications achieved $p < 0.005$. C-indexes were 0.741 (95% CI 0.690–0.799) and 0.727 (95% CI 0.676–0.841) for the RFA nomogram, as well as 0.789 (95% CI 0.744–0.845) and 0.719 (95% CI 0.642–0.820) for the SR nomogram, in training and validation cohorts, respectively.

Clinical Use

The decision curve analysis for the RFA nomogram only using clinical variables (age, PLT, and tumor size) and that with Radiomics signatures integrated are demonstrated in Figure 4e (green and red curves). It showed that if the threshold probability of a patient was $>30\%$, integrating Radiomics signatures and clinical variables to predict his/her 2-year PFS after receiving RFA adds remarkably more net benefit than only using clinical variables. We further applied the same analysis for the SR nomogram (Fig. 4f), and Radiomics signatures also provided more net benefit to predict the 2-year PFS of a patient receiving SR than only using clinical variables, if his/her threshold probability was $>15\%$.

Optimization of Treatment Selection between RFA and SR

After we established 2 DL-based Radiomics models (R-RFA and R-SR) that achieved effective prognostic stratification in each corresponding patient group (RFA and SR), we swapped the application of these 2 models into their opposite groups for restratifications. Mann-Whitney U tests proved that the second stratifications in both patient groups also achieved significant differences in hazards (both $p < 0.001$), indicating that R-RFA and R-SR effectively stratified SR and RFA into low- and high-risk subgroups. In RFA, we identified 151 patients stayed in the same subgroups (130 in low-risk and 21 in high-risk) regardless of using R-RFA or R-SR for prognostic stratifications (Fig. 5a). However, 63 patients changed their categories, and 37 of them originally considered to be high-risk in receiving RFA were suggested to be low-risk in receiving SR (Fig. 5a, red category), which was 63.8% patients in the original high-risk subgroup and 17.3% patients in the entire RFA group.

Then, we applied the RFA nomogram and SR nomogram to calculate the individualized 2-year PFS probability of these 37 patients, respectively. The average probability was increased from 0.62 ± 0.10 to 0.74 ± 0.01 (Fig. 5b), which indicated a significant rise of expectation ($p < 0.001$) for achieving 2-year PFS, if these patients were treated by SR instead of RFA.

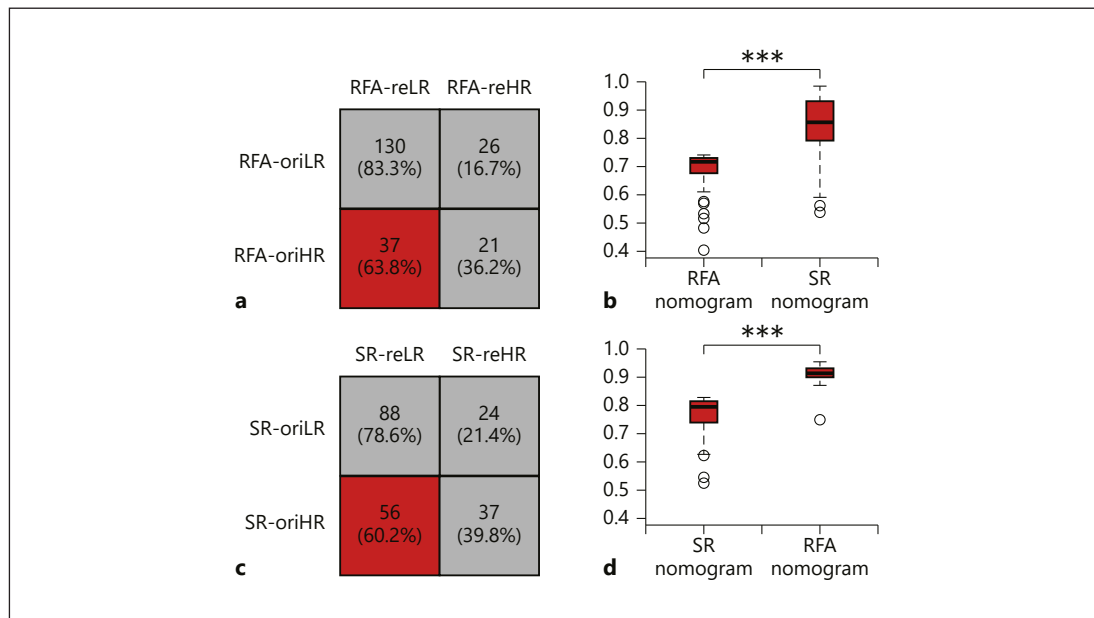


Fig. 5. Comparisons between 2 stratifications using 2 Radiomics models. **a** In the RFA group, the original stratification given by R-RFA and the reconstituted stratification given by R-SR generated 4 subgroups. Thirty-seven patients (red category) stratified as high-risk for receiving RFA were considered low-risk for receiving SR. **b** For these 37 patients, distributions of their predicted probability for 2-year PFS given by RFA and SR nomograms were plotted and compared. **c** In the SR group, the original and reconstituted stratifications given by R-SR and R-RFA also generated for subgroups. Fifty-six patients (red category) stratified as high-risk for receiving SR were considered low-risk for receiving RFA. **d** For these 56 patients, distributions of their predicted probability for 2-year PFS given by RFA and SR nomograms were plotted and compared. *** $p < 0.001$. RFA, radiofrequency ablation; SR, surgical resection.

Similarly, in SR, we identified 56 patients originally considered to be high-risk in receiving SR were suggested to be low-risk in receiving RFA (Fig. 5c, red category), which was 60.2% patients in the original high-risk subgroup and 27.3% patients in the entire SR group. The nomogram analysis revealed that the average probability of them for achieving 2-year PFS was increased from 0.76 ± 0.12 to 0.91 ± 0.05 (Fig. 5d, $p < 0.001$).

After switching the DL-based algorithm, there were patients whose subgroup was inconsistent with the original high-risk subgroup (37 for RFA group, and 56 for SR group). There were also patients whose subgroup was consistent with the original high-risk subgroup (21 for RFA group and 37 for SR group). We applied statistical analysis about the clinical characteristics for the high-risk patients showing inconsistency and consistency (online suppl. Table S6). The patients in low-risk subgroups means they obtained proper treatment, so we did not compare their clinical characteristics.

Discussion

In the current phase, the main applications of DL on US diagnosis are focused on liver, thyroid, and breast. B-mode images, shear-wave elastography (SWE), and CEUS were the most common kind of ultrasound used in literatures. A DL model proposed by Wang et al. [27] can assess the liver fibrosis in two-dimensional SWE. The presence of metastatic lymph nodes for thyroid carcinomas was also proved to be predictable by a well-designed DL-based CAD

system [28]. For the classification of breast tumors with SWE, it is proved that DL-based method can achieve significantly higher prediction performance than other prediction algorithms using human-defined features [29]. There was a study showing that the DL model was able to predict benign or malignancy for focal liver lesions using CEUS cines [30]. However, temporal features extracted from time intensity curves of CEUS cines were still human-defined rather than automatically learned by the DL model in that study. To the best of our knowledge, the development of DL models for automatic analysis of CEUS cines, as well as applying such approach for prognosis prediction and treatment selection between RFA and SR, has not yet been reported.

In this study, we developed and validated 2 DL-based Radiomics models (R-RFA and R-SR) for preoperative predictions of PFS in very-early or early-stage HCC patients received RFA and SR. Both models quantitatively interpreted the heterogeneity of HCC dynamic behavior by intelligently extracting and learning numerous spatiotemporal features from CEUS cines. Then, they successfully constructed Radiomics signatures (hazard values) and correlated them with PFS predictions for different treatments. As a result, R-RFA- and R-SR-stratified RFA and SR patient groups into low- and high-risk subgroups with remarkable accuracy. In RFA, the C-index researched 0.754 and 0.726 in training and validation cohorts, and in SR, the C-indexes were 0.787 and 0.741. There was no significant difference for patients examined by different US vender (online suppl. Table S5). We further applied them to predict the 2-year PFS for RFA and SR groups. AUCs were 0.820 and 0.815 in training and validation cohorts for RFA, as well as 0.863 and 0.828 for SR (online suppl. Fig. S4). These results further proved the effectiveness of the 2 DL-based Radiomics models for prognosis predictions of RFA and SR.

As the DL process is generally considered as a “black box” for analyzing big data [13], we transferred CEUS cines into dynamic Radiomics feature maps coded with pseudo-colors, in order to visually observe how AI would interpret these ultrasound videos for prognosis prediction. By observing network visualization results, we found that the R-RFA focused on the entire tumor is during arterial and portal phases, whereas R-SR paid more attention to the tumor periphery during all 3 phases. We associated this phenomenon with the surgical procedures and basic physics of RFA and SR. The operation of RFA is based on the heat transfer from the tip of the probe to entire HCC to ablate tumor cells, whereas SR resects the tumor tissue along its margin and physically separates the tumor with its surrounding parenchyma, so that the tumor central area is far less important than its boundary. We could make analogy between this and DL-based face recognition algorithms. Previous studies have revealed that DL models tend to pay more attention to the specific face areas such as human eyes, noses, and mouth to predict the input’s identity, because these areas have more details related with human identity [31, 32]. We think our DL models also intelligently found the critical areas and dynamic phases with more details corresponding with the treatment outcomes for RFA and SR. This partially revealed the mechanism of why our DL approaches successfully achieved prediction and also indicated that the DL models were effectively trained. This indicated that the 2 DL-based models were well established to interpret CEUS cines with regard to different treatments, and the visualization of their feature maps may benefit the human observation and judgment in clinical practices.

After the multivariable Cox analysis of clinical variables, we identified age, PLT, and tumor size were significant risk factors for RFA, whereas ALT, ALB, and tumor size were significant for SR. Therefore, both liver dysfunction and tumor size were prognostic factors for RFA and SR, which were consistent with other studies [6–8]. However, comparing with these clinical variables, Radiomics signatures built by R-RFA and R-SR made a dominant contribution in predicting PFS for RFA and SR. To predict the 2-year PFS for each individual [33, 34], we established the RFA nomogram and SR nomogram by incorporating their corre-

sponding Radiomics signatures and significant clinical variables. Then, excellent calibrations between prediction and observation were confirmed for both nomograms, and they also achieved satisfactory discrimination in each treatment group (In RFA, C-indexes, 0.741 and 0.727, in training and validation cohorts; In SR, C-indexes, 0.789 and 0.719, in training and validation cohorts).

Comparing with R-RFA and R-SR, RFA and SR nomograms showed very limited improvement of accuracy for the individualized 2-year PFS prediction. In RFA, AUCs increased 0.011 and 0.007 in training and validation cohorts. In SR, AUCs increased 0.034 and 0.013 in training and validation cohorts. Furthermore, after applying the decision curve analysis, we found that using the integrated Radiomics signature and clinical variables to predict the 2-year PFS for RFA and SR brought much more benefit than using clinical variables alone. All these results revealed that the CEUS examination and DL-based Radiomics analysis can be used to facilitate the preoperative risk stratification and individualized PFS prediction for very-early or early-stage HCC patients scheduled for RFA or SR treatment. Because applying the DL-based Radiomics method only requires very little manual work for defining the initial ROI, this approach is extremely easy-to-use and can be directly integrated into the routine CEUS work-flow.

Finally, we applied both R-RFA and R-SR to all enrolled 419 patients, so that each individual was stratified twice regarding to different treatments. Our results revealed that 17.3% patients ($n = 37$) in the RFA group and 27.3% patients ($n = 56$) in the SR group should change their original treatment into the other. If they had done that, their average probability of 2-year PFS would increase 12 and 15%, respectively, which were calculated by our RFA and SR nomograms. Both improvements were statistically significant (both $p < 0.001$). These results suggested that the DL-based Radiomics analysis of CEUS cines was likely to be a new and effective tool for optimized treatment selection between RFA and SR in patients with very-early or early-stage HCC. This novel approach holds a great potential to compensate other proposed methods designed for choosing RFA or SR [6–8] and improve the overall clinical decision making for HCC patients. In addition, no significant difference was found on the clinical characteristics for patients who need to swap their treatments. This proved that the optimization of HCC treatments is a complex work which cannot be determined by certain clinical characteristics. Therefore, analyzing the heterogeneity of HCC dynamic behavior recorded in CEUS cines is essential for the optimized treatment selection between RFA and SR for patients with very-early or early-stage HCC.

Our study has 2 major limitations. The study population was limited in size and was retrospectively acquired from a single-center hospital. The DL-based model lacks interpretability and behaved in a “black-box” manner, which is an inherent pitfall in the field of AI. Therefore, a multi-center perspective study with a larger patient population is still necessary in the future to further validate the performance of our Radiomics models and nomograms. Our DL network also needs to be further optimized with better engineering design to continuously improve its overall performance and interpretability.

In conclusion, this study demonstrated 2 DL-based Radiomics models and nomograms that could be conveniently applied to preoperative CEUS examinations, and they could facilitate the accurate PFS prediction and optimized treatment selection of RFA and SR for patients with very-early or early-stage HCC.

Acknowledgements

We thank all participants for their endeavor and contribution in this study. We thank the First Affiliated Hospital of Sun Yat-sen University and the Institute of Automation Chinese Academy of Sciences for their support of this research project.

Statement of Ethics

The patients in this study have given their written informed consent, and the study protocol was approved by the institute's committee on human research.

Disclosure Statement

The authors have no conflicts of interest to declare.

Funding Sources

This work was supported by grants from Ministry of Science and Technology of China under Grant (Nos. 2017YFA0205200 and 2016YFC0103803), the State Key Project on Infectious Diseases of China (2018ZX10723204), National Natural Science Foundation of China under Grant (Nos. 61671449, 81227901, 81527805, and 81530055), Chinese Academy of Sciences under Grant (No. GJJSTD20170004, KFJ-STZ-ZDTP-059, YJKYYQ20180048, XDBS01030200, and QYZDJ-SSW-JSC005), and the Science and Technology Development Special Fund of Guangdong Province (2017A020215011). The funders had no role in study design, data collection, data analysis, interpretation, writing of the report, and the decision of submitting it for publication.

Author Contributions

K.W., M.L., X.X., and J.T.: study conception and design. F.L., D.L., X.X., L.S., M.K., G.H., B.P., and Y.W.: data collection. F.L., D.L., K.W., and M.L.: data analysis. F.L., D.L., K.W., and M.L.: manuscript drafting. All authors: final approval of the manuscript.

References

- 1 Galle PR, Forner A, Llovet JM, Mazzaferro V, Piscaglia F, Raoul JL, et al.; European Association for the Study of the Liver. Electronic address: easloffice@easloffice.eu; European Association for the Study of the Liver. Management of hepatocellular carcinoma. *J Hepatol*. 2018 Jul;69(1):182–236.
- 2 Chen MS, Li JQ, Zheng Y, Guo RP, Liang HH, Zhang YQ, et al. A prospective randomized trial comparing percutaneous local ablative therapy and partial hepatectomy for small hepatocellular carcinoma. *Ann Surg*. 2006 Mar;243(3):321–8.
- 3 Huang J, Yan L, Cheng Z, Wu H, Du L, Wang J, et al. A randomized trial comparing radiofrequency ablation and surgical resection for HCC conforming to the Milan criteria. *Ann Surg*. 2010 Dec;252(6):903–12.
- 4 Feng K, Yan J, Li X, Xia F, Ma K, Wang S, et al. A randomized controlled trial of radiofrequency ablation and surgical resection in the treatment of small hepatocellular carcinoma. *J Hepatol*. 2012 Oct;57(4):794–802.
- 5 Bruix J, Sherman M; Practice Guidelines Committee, American Association for the Study of Liver Diseases. Management of hepatocellular carcinoma. *Hepatology*. 2005 Nov;42(5):1208–36.
- 6 Cucchetti A, Piscaglia F, Cescon M, Serra C, Colecchia A, Maroni L, et al. An explorative data-analysis to support the choice between hepatic resection and radiofrequency ablation in the treatment of hepatocellular carcinoma. *Dig Liver Dis*. 2014 Mar;46(3):257–63.
- 7 Cucchetti A, Piscaglia F, Cescon M, Colecchia A, Ercolani G, Bolondi L, et al. Cost-effectiveness of hepatic resection versus percutaneous radiofrequency ablation for early hepatocellular carcinoma. *J Hepatol*. 2013 Aug;59(2):300–7.
- 8 Leoni S, Piscaglia F, Serio I, Terzi E, Pettinari I, Croci L, et al. Adherence to AASLD guidelines for the treatment of hepatocellular carcinoma in clinical practice: experience of the Bologna Liver Oncology Group. *Dig Liver Dis*. 2014 Jun;46(6):549–55.
- 9 Alzaraa A, Gravante G, Chung WY, Al-Leswas D, Morgan B, Dennison A, et al. Contrast-enhanced ultrasound in the preoperative, intraoperative and postoperative assessment of liver lesions. *Hepatol Res*. 2013 Aug;43(8):809–19.
- 10 Sato K, Tanaka S, Mitsunori Y, Mogushi K, Yasen M, Aihara A, et al. Contrast-enhanced intraoperative ultrasonography for vascular imaging of hepatocellular carcinoma: clinical and biological significance. *Hepatology*. 2013 Apr;57(4):1436–47.

- 11 Maruyama H, Takahashi M, Shimada T, Sekimoto T, Kamesaki H, Kanai F, et al. Pretreatment microbubble-induced enhancement in hepatocellular carcinoma predicts intrahepatic distant recurrence after radiofrequency ablation. *AJR Am J Roentgenol*. 2013 Mar;200(3):570–7.
- 12 Zou RH, Lin QG, Huang W, Li XL, Cao Y, Zhang J, et al. Quantitative contrast-enhanced ultrasonic imaging reflects microvascularization in hepatocellular carcinoma and prognosis after resection. *Ultrasound Med Biol*. 2015 Oct;41(10):2621–30.
- 13 Wang S, Liu Z, Rong Y, Zhou B, Bai Y, Wei W, et al. Deep learning provides a new computed tomography-based prognostic biomarker for recurrence prediction in high-grade serous ovarian cancer. *Radiother Oncol*. 2019 Mar;132:171–7.
- 14 Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage*. 2006 Jul;31(3):1116–28.
- 15 Luck M, Sylvain T, Cardinal H, Lodi A, Bengio Y. *Deep learning for patient-specific kidney graft survival analysis*. NeuroIPS; 2017.
- 16 Breslow NE. Analysis of survival data under the proportional hazards model. *Int Stat Rev*. 1975;43(1):45–57.
- 17 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 Nov;10(21):7252–9.
- 18 Dignam JJ, Zhang Q, Kochevinsky M. The use and interpretation of competing risks regression models. *Clin Cancer Res*. 2012 Apr;18(8):2301–8.
- 19 Ueno H, Mochizuki H, Akagi Y, Kusumi T, Yamada K, Ikegami M, et al. Optimal colorectal cancer staging criteria in TNM classification. *J Clin Oncol*. 2012 May;30(13):1519–26.
- 20 Stish BJ, Pisansky TM, Harmsen WS, Davis BJ, Tzou KS, Choo R, et al. Improved metastasis-free and survival outcomes with early salvage radiotherapy in men with detectable prostate-specific antigen after prostatectomy for prostate cancer. *J Clin Oncol*. 2016 Nov;34(32):3864–71.
- 21 Chan AW, Zhong J, Berhane S, Toyoda H, Cucchetti A, Shi K, et al. Development of pre and post-operative models to predict early recurrence of hepatocellular carcinoma after surgical resection. *J Hepatol*. 2018 Dec;69(6):1284–93.
- 22 Johnson PJ, Berhane S, Kagebayashi C, Satomura S, Teng M, Reeves HL, et al. Assessment of liver function in patients with hepatocellular carcinoma: a new evidence-based approach—the ALBI grade. *J Clin Oncol*. 2015 Feb;33(6):550–8.
- 23 Song J, Shi J, Dong D, Fang M, Zhong W, Wang K, et al. A New Approach to Predict Progression-free Survival in Stage IV EGFR-mutant NSCLC Patients with EGFR-TKI Therapy. *Clin Cancer Res*. 2018 Aug;24(15):3583–92.
- 24 Liu Z, Zhang XY, Shi YJ, Wang L, Zhu HT, Tang Z, et al. Radiomics analysis for evaluation of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *Clin Cancer Res*. 2017 Dec;23(23):7253–62.
- 25 Kramer AA, Zimmerman JE. Assessing the calibration of mortality benchmarks in critical care: the Hosmer-Lemeshow test revisited. *Crit Care Med*. 2007 Sep;35(9):2052–6.
- 26 Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*. 2006 Nov-Dec;26(6):565–74.
- 27 Wang K, Lu X, Zhou H, Gao Y, Zheng J, Tong M, et al. Deep learning Radiomics of shear wave elastography significantly improved diagnostic performance for assessing liver fibrosis in chronic hepatitis B: a prospective multicentre study. *Gut*. 2019 Apr;68(4):729–41.
- 28 Lee JH, Baek JH, Kim JH, Shim WH, Chung SR, Choi YJ, et al. Deep learning-based computer-aided diagnosis system for localization and diagnosis of metastatic lymph nodes on ultrasound: a pilot study. *Thyroid*. 2018 Oct;28(10):1332–8.
- 29 Zhang Q, Xiao Y, Dai W, Suo J, Wang C, Shi J, et al. Deep learning based classification of breast tumors with shear-wave elastography. *Ultrasonics*. 2016 Dec;72:150–7.
- 30 Wu K, Chen X, Ding M. Deep learning based classification of focal liver lesions with contrast-enhanced ultrasound. *Optik (Stuttg)*. 2014;125(15):4057–63.
- 31 Taigman Y, Yang M, Ranzato MA, Wolf L. DeepFace: closing the gap to human-level performance in face verification. 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH; 2014. p. 1701–8.
- 32 Masi I, Tran AT, Leksut JT, Hassner T, Medioni G. *Do we really need to collect millions of faces for effective face recognition?* CVPR; 2016.
- 33 Imamura H, Matsuyama Y, Tanaka E, Ohkubo T, Hasegawa K, Miyagawa S, et al. Risk factors contributing to early and late phase intrahepatic recurrence of hepatocellular carcinoma after hepatectomy. *J Hepatol*. 2003 Feb;38(2):200–7.
- 34 Portolani N, Coniglio A, Ghidoni S, Giovanelli M, Benetti A, Tiberio GA, et al. Early and late recurrence after liver resection for hepatocellular carcinoma: prognostic and therapeutic implications. *Ann Surg*. 2006 Feb;243(2):229–35.