**Deep learning Radiomics based on contrast-enhanced ultrasound might optimize curative treatments for very early or early stage hepatocellular carcinoma patients**

1. **Supplementary methods**
2. **Supplementary results**
3. **Supplementary references**
4. **Supplementary tables**
5. **Supplementary figures**

**Supplementary methods**

*Concepts of Convolutional Neural Network (CNN)-based Cox proportional hazards regression (Cox-CNN)*

We combined the CNN model and Cox proportional hazards regression algorithm to predict the PFS for RFA and LR in this study. In general, Cox-CNN comprised five layers: 2D convolution, 2D max-pooling, global aggregation, fully connected, and cox-loss function. The five layers and the training method of Cox-CNN are described below.

*2D convolution*

The convolution operation is a common method used to extract image features with human-defined or auto-learning filters. In CNNs, convolutional filters are learned by automatically analyzing a number of samples [1]. The mathematical definition of 2D convolution is as follows:



where  is the inputted CEUS cines,  is the 2D convolutional filter, and  is the convolutional result, which is the CEUS features at one layer of the CNN model.

In this study, we applied the 2D convolution operation on all the frames and left the correlation between the frames unprocessed until the feature flow reached the global aggregation module. Because, compared with 2D convolution, 3D convolution, which processes spatio-temporal features simultaneously, possesses more parameters and is prone to overfitting, we chose 2D depthwise-convolution [2] to further reduce the number of parameters while maintaining the auto-learning ability of the convolution operation. The ReLU was used as the activation function of the models [3].

*2D max-pooling*

The pooling operation was applied to aggregate information locally in one CEUS frame [1]. Max-pooling specifically uses the maximum value to represent one neighborhood area. By stacking convolution and the max-pooling operation, the high level of CNN could have a perception in the entire field of one CEUS frame. In this study, we set the pool neighborhood area to 2-by-2.

*Global aggregation*

The stacking of 2D convolution and the max-pooling operation aggregated the information in the level of the frame. Then, the global aggregation module was applied to learn the information along the time dimension in CEUS cines. For simplicity and efficacy, the neural network variant of the vector of locally aggregated descriptors (NetVLAD) [4-7] was selected to learn genuine spatio-temporal CEUS features. See Fig. S1 for the detailed structure of the NetVLAD module.

*Fully connected*

Depending on specific problems, the fully connected layer can be used as a classifier or a regressor [1]. In this study, with the spatio-temporal features of CEUS cines after global aggregation, the fully connected layer performed regression computation to obtain the corresponding survival hazard value.

*Cox-loss function*

A loss function is needed to guide the CNN model to optimize its parameters to generate more accurate predictions. We adopted the Cox framework [8, 9] to do survival analysis in this study. The model regresses a hazard value for each sample. The hazard value was a real number without pre-defined range. It is expected that the patient with shorter PFS time would be assigned to a higher hazard value. Because there is not a fixed scale for predicted hazard values, instead of the actual value itself, it is the mutual sequence order of hazard values that matters. In our case, the scale of hazard values was [-1.5, 4.5]. The Cox proportional hazard model has been used as a survival model in most previous studies [8, 10, 11]. However, there exists several limitations in the Cox model; for example, the input of the common Cox model must be well-defined features instead of raw data such as CEUS cines, and only the linear combination of features is implemented [9]. To overcome the drawbacks of the common Cox model, we propose to automatically learn features using CNN models, and define the loss function of the CNN as the Cox log partial likelihood function. The definition of the loss function is as follows:



where we used  to represent the hazard values regressed by the network with model’s parameters , and . In survival analysis, if a patient did not have the observation event (radiological identification of tumor progression in our study) during the observation time (from January 2008 to January 2018 in our study), the patient was regarded as censored. We used  to indicate that the patient  is censored. And instead,  meant not censored. We used  to represent the duration from the date RFA or SR was performed to the radiological identification of tumor progression or the end of observation time. From experiments, we found that the more samples inputted into the model simultaneously, the better the model performance achieves.

In the loss function of Cox model, what the labels  provided were the sequence order of patients by the survival time. Therefore, the learning target of our Cox-CNN model was to learn the sequence order between different patients, instead of regressing the actual survival time length directly. We tried to regress hazard values to make the hazard values of patients who lived shorter higher than the hazard values of patients who lived longer. The survival time length of a specific patient can be computed by post-processing based on hazard values [12].

*Structure of Cox-CNN*

In this study, we designed a custom DL network based on the specific application for dynamic contrast-enhanced ultrasound (CEUS). As it is common practice to split the CEUS cines into arterial phase (AP), portal phase (PP), and venous phase (VP) in clinical practice, we chose to automatically learn the spatio-temporal features for each phase separately and then globally aggregate the three local features into one global feature via the global aggregation layer. The characteristic of the CEUS cines could be represented by global feature, as shown in the section “Visualization of DL-based Radiomics models”.

We used the same network structure for the survival analysis of RFA and SR. The Cox-CNN consisted of four CNN modules, followed by one flatten layer and one global aggregation module (NetVLAD) [4-7]. Each CNN module was composed of a convolution layer, batchnorm layer[13], ReLU function, and max-pooling layer. We adopted 2D convolution operation to learn features for each frame separately. The convolution network adopted depthwise convolution operation [3] for reducing the number of parameters, while preserving similar learning capability as common convolution (Table S5). The convolution network was shared by AP, PP, and VP. After the last convolution block, we global-averagely pooled the features along spatial dimensions, so each frame was compressed as a vector with the size of 96 in our experiments. Then the pooled tensor was processed by a VLAD module [4] to model the temporal correlation among different frames. The output of VLAD was a vector whose size was 64. Finally, we used a fully connection layer to regress the hazard values based on the 64-length vector. We provided the detailed structures of each part in the Table S4. Survival hazard was regressed as the output of Cox-CNN, and we implemented a loss function based on the mathematical definition of partial likelihood, which is also the target function of Cox proportional hazards. Notably, we achieved an end-to-end survival analysis neural network. In contrast to most of the other works[9, 11], our proposed model could auto-learn features directly from CEUS, instead of requiring human-defined features. See Fig. 2b for the network structure of Cox-CNN.

We initialized the parameters of the last fully connection layer by a uniform distribution whose range was from -0.01 to 0.01. Other layers’ parameters were initialized by the PyTorch’s default initialization setting.

*Training method of Cox-CNN*

We used Adam [14] as the optimizer with a batch size of 128. As the network need to learn a progression time order using cox-loss function, a large batch size is essential in training Cox-CNN. With the learning rate of 0.0001, we trained all the models in 50 epochs. Data augmentation including random rotation, flip, and crop was applied to reduce the risk of overfitting. Center crop was used when validating the model. For the input data, the height and width of tumor region of interest (ROI) from each frame were 64 pixels. The time durations of AP, PP, and VP were different. We resampled and unified the frame rate of original CEUS cines to 2 frames per second (fps) in the course of preprocessing, thus every high frame rate cine (> 20 fps) was resampled into 10 low frame rate cines without overlapping. As it is clinically common to record the microcirculation perfusion of HCC from the 8th second, we adopted the CEUS cines from 8th second to 30th second as AP, so for a single input of the DL model, the size of AP was 44×64×64. Similarly, we used the PP frames from the 30th second to the 70th second, and the size of PP was 80×64×64. The used VP was from the 120th second to the 138th second, and the size of VP was 36×64×64. We learned the entire network at once for DL network learning.

We adopted 5-fold cross validation method to search for best hyper-parameters setting. In detail, we firstly divided our dataset into training and validation cohorts by the way described in Materials and Methods section, Development and validation of DL-based Radiomics models subsection. Then, we divided training cohort into 5 nonoverlapping parts. When we tested the effectiveness of one set of hyper-parameters, we used successively one part to validate the prediction performance (C-index in our case) and the rest as training dataset to train our network. Finally the C-index values in 5 times of training were averaged as the final performance for the set of hyper-parameters. In this way, we search for a suitable set of hyper-parameters and an effective network structure. With the determined hyper-parameters and network structure, we trained the model using the entire training cohort, and estimated the generalized prediction performance in the prepared validation cohort.

*Measuring the accuracy of Radiomics and Individualized models*

The prediction accuracy of models were assessed by time-dependent receiver operating characteristic (ROC) analysis [15]. Two-year PFS ROC curves were plotted and the area under curves (AUC) was quantified for DL-based Radiomics and individualized models in the training and validation cohorts.

*Visualization of DL-based Radiomics models*

To further explore how DL-based Radiomics models interpret CEUS cines for prognostic prediction, we transformed Radiomics features of R-RFA and R-SR into pseudo-colored maps by Selvaraju’s method [16]. Features characteristics learned automatically by R-RFA and R-SR were visualized by the distribution of different colors. In transformed pseudo-colored CEUS cines, pixels with warmer colors (e.g., red and yellow) indicated stronger correlation with the prognostic prediction than pixels with colder colors (e.g., blue and green).

**Supplementary results**

*Complications*

There were no treatment-related death in SR and RFA groups. Major complications were significantly more frequent in SR group than in RFA group (32/205 vs 5/214, *P* <0.01). In SR group, major complications included liver failure (n = 1), liver abscess (n = 1), choledochal stenosis (n = 1), acute cholecystitis (n = 1), persistent jaundice for more than 30 days after surgery (n = 3), moderate/severe ascites (n = 15) and pleural effusion needing percutaneous drainage (n = 10). In RFA group, major complications were as follows: tumor seeding (n = 1), biloma (n = 2), abdominal bleeding (n = 2). Minor complications in SR and RFA groups included fever of higher 38.5◦C (15/205 vs 6/214, P = 0.04) and pain requiring analgesics after treatment (65/205 vs 12/214, P < 0.01). The complications were also presented as early and late complications based on the occurrence time post-treatment (< 1 month and ≥ 1 month). In SR group, early complications included liver failure (n = 1), acute cholecystitis (n = 1), persistent jaundice for more than 30 days after surgery (n = 3), moderate/severe ascites (n = 15), pleural effusion needing percutaneous drainage (n = 10), fever (n = 15), and pain (n = 65). Late complications in SR included liver abscess (n = 1) and choledochal stenosis (n = 1). In RFA group, early complications were as follows: abdominal bleeding (n = 2), fever (n = 6) and pain (n = 12). Late complications in RFA included tumor seeding (n = 1) and biloma (n = 2).

*Measuring the accuracy of Radiomics and Individualized models*

We applied R-RFA and R-SR to predict the two-year PFS for patients in RFA and SR. AUCs of training and validation cohorts reached 0.820 (CI: 0.742-0.889) and 0.815 (CI: 0.717-0.923) for R-RFA, as well as 0.863 (CI: 0.795-0.921) and 0.828 (CI: 0.719-0.899) for R-SR. The corresponding ROC curves for R-RFA and R-SR are presented in Fig. S4a-b.

AUCs of ROC curves were 0.831 (CI: 0.758-0.903) and 0.822 (CI: 0.719-0.925) for the RFA nomogram, as well as 0.897 (CI: 0.842-0.951) and 0.841 (CI: 0.718-0.964) for the SR nomogram, in training and validation cohorts, respectively (Fig. S4c-d).

*The influence of different US venders*

The data used in our study was collected from two kinds of US venders: Philips iU22 and Toshiba Aplio. The numbers of CEUS cines obtained using Philips iU22 and Toshiba Aplio were 67 and 147 for the RFA group, 75 and 130 for the SR group, respectively. Hazards values were computed for all patients in RFA and SR groups, using R-RFA and R-SR models respectively. We compared the statistical difference of hazards between samples of two kinds of US systems using Mann-Whitney U test. The *P* values were shown in Table S5. There was no significant difference for the output between two kinds of US, neither in training cohorts nor validation cohorts (all *P* > 0.1). Therefore, these two different US venders had no impact on the output of our DL models.

*Comparison of clinical backgrounds and network features for patients in different subgroups*

We compared the clinical backgrounds for patients in inconsistency and consistency subgroups. The results were shown in the Table S6.

We extracted the features of Cox Regression layer. In order to show the features distribution, principal component analysis (PCA) was adopted to reduce the dimension of the features from 64 to 2. We plotted the features distribution map using bivariate kernel density estimation method. With the color changing from blue to red, the density increased smoothly. We marked the location of patients of low-/high-risk subgroups (Fig. S5 a, b) and inconsistency/consistency subgroups (Fig. S5 c, d) in this 2-D features space. It could be observed that although with some outliers, the patients in one subgroup tend to be clustered. Besides, the patients in low-risk and patients in inconsistency subgroups tend to be closer to the region with higher features distribution density. The observation result revealed that there was a hierarchy in the network feature distribution space for patients with different kinds of survival outcomes. Patients who might be benefited from switching treatment could be identified by corresponding network models and tend to be encoded by network into similar features.

**References**

[1] LeCun Y, Bottou L, Bengio Y, Haffner P: Gradient-based learning applied to document recognition. Proceedings of the IEEE 1998;86:278-324.

[2] Francois C: Xception: Deep learning with depthwise separable convolutions. CVPR 2016;1251-1258.

[3] Glorot X, Bordes A, Bengio Y: Deep sparse rectifier networks. AISTATS 2011;315-323.

[4] Tolias G, Avrithis Y, Jegou H: Image search with selective match kernels: Aggregation across single and multiple images. IJCV 2016;116:247-261.

[5] Arandjelovic R, Gronat P, Torii A, Pajdla T, Sivic J: NetVLAD: CNN architecture for weakly supervised place recognition. TPAMI 2016;5297-307.

[6] Girdhar R, Ramanan D, Gupta A, Sivic J, Russell B, et al: ActionVLAD: Learning spatio-temporal aggregation for action classification. CVPR 2017;971-980.

[7] Lin R, Xiao J, Fan J: NeXtVLAD: An efficient neural network to aggregate frame-level features for large-scale video classification. ECCV 2018;206-218.

[8] Breslow NE: Analysis of survival data under the proportional hazards model. ISI 1975;43:45-57.

[9] Luck M, Sylvain T, Cardinal H, Lodi A, Bengio Y: Deep learning for patient-specific kidney graft survival analysis. NeurIPS 2017.

[10] Zhang B, Tian J, Dong D, Gu D, Dong Y, Zhang L, et al: Radiomics features of multiparametric MRI as novel prognostic factors in advanced nasopharyngeal carcinoma. Clin Cancer Res 2017;23:4259-4269.

[11] Tangri N, Stevens LA, Griffith J, Tighiouart H, Djurdjev O, Naimark D, et al: A predictive model for progression of chronic kidney disease to kidney failure. JAMA 2011;305:1553-1559.

[12] Cox D. Regression models and life tables. J Roy Stat Soc B 1972:34:187-220.

[13] Sergey I, Christian S: Batch Normalization: Accelerating deep network training by reducing internal covariate shift. ICML 2015;448-456.

[14] Kingma DP, Ba J: Adam: A method for stochastic optimization. ICLR 2015.

[15] 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 EGFR0mutant NSCLC patients with EGFR-TKI therapy. Clin Cancer Res 2018;24:3583-3592.

[16] Selvaraju RR, Cogswell M, Das A, Ramakrishna V, Parikh D, Batra D: Grad-CAM: Visual explanations from deep networks via gradient-based localization.ICCV 2017;618-626.

**Table S1.** Baseline characteristics in the low- and high-risk subgroups for training and validation cohorts

| **Characteristic** | **RFA (n=214 )** | | | **SR (n= 205)** | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Low risk (n=156)** | **High risk (n=58)** | ***P*** | **Low risk (n=112)** | **High risk (n=93)** | ***P*** |
| **Age (years) (mean±SD)** | 56.64±11.38  ( 29– 83 ) | 55.05±9.86  (33-74) | 0.348 | 54.80±11.52  (20–78) | 53.03±10.61  (26-72) | 0.258 |
| ≤60 | 97 (62.18%) | 41 (70.69%) |  | 75 (66.96%) | 66 (70.97%) |  |
| >60 | 59 (37.82%) | 17 (29.31%) |  | 37 (33.04%) | 27 (29.03%) |  |
| **Sex** |  |  | 0.471 |  |  | 0.998 |
| Male | 140 (89.74%) | 50 (86.21%) |  | 94 (83.93%) | 79 (84.95%) |  |
| Female | 16 (10.26%) | 8 (13.79%) |  | 18 (16.07%) | 14 (15.05%) |  |
| **Liver cirrhosis** |  |  | 0.758 |  |  | 0.985 |
| Yes | 87 (55.77%) | 34 (58.62%) |  | 54 (48.21%) | 45 (48.39%) |  |
| No | 69 (44.23%) | 24 (41.37%) |  | 58 (51.79%) | 48 (51.61%) |  |
| **Performance status** |  |  | 0.590 |  |  | 0.859 |
| 0 | 117 (75.1%) | 46 (79.3%) |  | 90 (80.2%) | 76 (81.7%) |  |
| 1 | 39 (22.8%) | 12 (20.7%) |  | 22 (19.8%) | 17 (18.3%) |  |
| **AFP, ng/mL** |  |  | 0.073 |  |  | 0.538 |
| <20 | 77 (49.36%) | 19 (32.76%) |  | 47 (41.96%) | 34 (36.56%) |  |
| 20-200 | 42 (26.92%) | 28 (48.27%) |  | 21 (18.75%) | 22 (23.66%) |  |
| ≥200 | 37 (23.71%) | 11 (18.97%) |  | 44 (39.29%) | 37 (39.78%) |  |
| **ALT, U/L** | 42.17±33.36  (10.0–224.0) | 40.60±33.72  (4.3-194.0) | 0.761 | 47.41±55.97  (12–411) | 53.87±48.33  (11-448) | 0.383 |
| **TBIL, µmol/L** | 17.80±9.56  (4.0–62.9 ) | 18.33±11.70  (6.1-72.2) | 0.735 | 16.22±9.24  (3.9–77.8) | 16.67±6.58 (7.1-37.5) | 0.694 |
| **PT, seconds** | 13.12±1.62  (10.9–23.2) | 13.06±1.21  (11.4-15.9) | 0.798 | 12.86±2.08  (10.7–31.0) | 12.79±1.03  (10.7-16.0) | 0.767 |
| **ALB, g/L** |  |  | 0.985 |  |  | 0.428 |
| <35 | 29 (18.59%) | 10 (17.24%) |  | 14 (12.5%) | 16 (17.20%) |  |
| ≥35 | 127 (81.41%) | 48 (82.76%) |  | 98 (87.5%) | 77 (82.80%) |  |
| **WBC(×109/L) (mean ± SD)** | 5.97±3.81  (1.19-46.0) | 5.25±1.83  (2.24-11.07) | 0.169 | 5.92±2.87  (1.18-29.48) | 5.95±2.76  (2.26-19.56) | 0.939 |
| **PLT(×109/L) (mean ± SD)** | 161.99±82.78 (17-468) | 134.12±49.36 (42-262) | 0.017 | 168.25±67.79 (19-482) | 163.29±60.06 (42-340) | 0.584 |
| **Tumor size (cm)** |  |  | 0.606 |  |  | 0.848 |
| ≤ 2 | 111 (71.15%) | 44 (75.86%) |  | 33 (29.46%) | 24 (25.81%) |  |
| 2 - 5 | 45 (28.85%) | 14 (24.14%) |  | 79 (70.54%) | 69 (74.20%) |  |
| **Tumor location** |  |  | 0.491 |  |  | 0.843 |
| Right lobe | 134 (85.90%) | 46 (79.31%) |  | 81 (72.32%) | 68 (73.12%) |  |
| Left lobe | 11 (7.05%) | 9 (15.52%) |  | 28 (25.00%) | 22 (23.65%) |  |
| Bilobar | 11 (7.05%) | 3 (5.17%) |  | 3 (2.68%) | 3 (3.23%) |  |
| **Perivascular location** |  |  | 0.086 |  |  | 0.856 |
| Yes | 37 (24.72%) | 7 (12.07%) |  | 21 (18.75%) | 16 (17.20%) |  |
| No | 119 (76.28%) | 51 (87.93%) |  | 91 (81.25%) | 77 (82.80%) |  |
| **Located in periphery of liver** |  |  | 0.164 |  |  | 0.555 |
| Yes | 45 (28.85%) | 11 (18.97%) |  | 36 (32.14%) | 34 (36.56%) |  |
| No | 111 (71.15%) | 47 (81.03%) |  | 76 (67.86%) | 59 (63.44%) |  |
| ALBI scores |  |  | 0.288 |  |  | 0.804 |
| 1 | 84 | 28 |  | 63 | 39 |  |
| 2 | 71 | 27 |  | 47 | 53 |  |
| 3 | 1 | 3 |  | 2 | 1 |  |

RFA, radiofrequency ablation; SR, surgical resection; 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.

**Table S2.** Radiomics signature and clinical characteristics that were in the individualized prediction model

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variables** | | **Model** | | |
| **β** | **HR (95%CI)** | ***P*** |
| **RFA** | **Age** | 0.702 | 2.017 (1.163-3.499) | 0.013 |
| **PLT** | 0.303 | 1.354 (0.764-2.400) | 0.026 |
| **tumor size** | 0.186 | 1.204 (0.569-2.549) | 0.045 |
| **Radiomics signature** | 1.831 | 6.240 (3.550-10.969) | <0.0001 |
| **SR** | **ALT** | 0.309 | 1.362 (0.814-2.278) | 0.039 |
| **ALB** | 0.807 | 2.242 (1.181-4.259) | 0.013 |
| **tumor size** | 1.012 | 2.752 (1.112-3.513) | 0.043 |
| **Radiomics signature** | 1.051 | 2.861 (2.178-3.756) | <0.0001 |

RFA, radiofrequency ablation; SR, surgical resection; PLT, platelet count; ALT, alanine aminotransferase; ALB, albumin; HR, hazard ratio; CI, confidence interval; 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.

**Table S5.** The statistical test results of hazard values between two Philips iU22 and Toshiba Aplio.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **RFA** | | | **SR** | | |
| **Philips iU22** | **Toshiba Aplio** | *P* | **Philips iU22** | **Toshiba Aplio** | *P* |
| T | -0.234±0.201 | -0.184±0.390 | 0.728 | 0.023±0.802 | -0.033±0.681 | 0.417 |
| V | -0.207±0.205 | -0.198±0.206 | 0.177 | -0.101±0.468 | -0.101±0.379 | 0.135 |

RFA, radiofrequency ablation; SR, surgical resection; T, training cohort; V, validation cohort.

**Table S6.** Baseline characteristics in the inconsistency and consistency subgroups

| **Characteristic** | **RFA** | | | **SR** | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Inconsistency (n=37)** | **Consistency (n=21)** | ***P*** | **Inconsistency (n=56)** | **Consistency (n=37)** | ***P*** |
| **Age (years) (mean±SD)** | 54.95±10.89  ( 33– 74 ) | 55.23±7.70  (40-66) | 0.917 | 54.11±10.31  (26–72) | 51.41±10.84  (34-70) | 0.229 |
| ≤60 | 26 (89.19%) | 15 (71.43%) |  | 38 (67.86%) | 28 (75.68%) |  |
| >60 | 4 (10.81%) | 6 (28.57%) |  | 18 (32.14%) | 9 (24.32%) |  |
| **Sex** |  |  | 0.443 |  |  | 0.847 |
| Male | 33 (89.19%) | 17 (80.95%) |  | 47 (83.93%) | 32 (86.49%) |  |
| Female | 4 (10.81%) | 4 (19.05%) |  | 9 (16.07%) | 5 (13.51%) |  |
| **Liver cirrhosis** |  |  | 0.821 |  |  | 0.403 |
| Yes | 22 (59.46%) | 12 (57.14%) |  | 25 (44.64%) | 20 (54.05%) |  |
| No | 15 (40.54%) | 9 (42.86%) |  | 31 (55.36%) | 17 (45.94%) |  |
| **Performance status** |  |  | 0.107 |  |  | 0.417 |
| 0 | 32 (86.5%) | 14 (66.7%) |  | 44 (78.6%) | 32 (86.5%) |  |
| 1 | 5 (13.5%) | 7 (33.3%) |  | 12 (21. 4%) | 5 (13.5%) |  |
| **AFP, ng/mL** |  |  | 0.214 |  |  |  |
| <20 | 12 (32.43%) | 7 (33.33%) |  | 17 (30.36%) | 17 (45.95%) |  |
| 20-200 | 18 (48.65%) | 10 (47.62%) |  | 14 (25.00%) | 8 21.62%) |  |
| ≥200 | 7 (18.92%) | 4 (19.05%) |  | 25 (44.64%) | 12 (32.43%) |  |
| **ALT, U/L** | 43.23±40.67  (4.3–194) | 35.95±13.88  (16-68) | 0.432 | 59.81±82.26  (11.0–396.0) | 60.43±69.16  (14-416) | 0.869 |
| **TBIL, µmol/L** | 17.05±9.62  (6.1–41.5 ) | 20.58±14.37  (7.3-72.2) | 0.267 | 15.84±5.95  (7.1–37.5) | 17.93±7.24 (8.6-34.9) | 0.132 |
| **PT, seconds** | 13.03±1.22  (11.4–15.9) | 13.11±1.16  (11.5-15.8) | 0.807 | 12.66±0.93  (10.9–15.7) | 12.85±1.06  (10.7-15.5) | 0.364 |
| **ALB, g/L** |  |  | 0.145 |  |  | 0.577 |
| <35 | 4 (10.81%) | 6 (28.57%) |  | 11 (19.64%) | 5 (13.51%) |  |
| ≥35 | 33 (89.19%) | 15 (71.43%) |  | 45 (80.36%) | 32 (86.49%) |  |
| **WBC(×109/L) (mean ± SD)** | 5.09±1.73  (2.24-9.47) | 5.53±1.93  (3.15-11.07) | 0.375 | 5.77±2.29  (2.56-17.7) | 6.22±3.32  (2.26-19.56) | 0.441 |
| **PLT(×109/L) (mean ± SD)** | 131.21±45.51 (61-253) | 139.23±55.13 (42-262) | 0.552 | 163.29±58.31 (47-340) | 163.28±62.62 (42-277) | 0.954 |
| **Tumor size (cm)** |  |  | 0.750 |  |  | 0.238 |
| ≤ 2 | 29 (78.38%) | 15 (71.43%) |  | 17 (30.36%) | 7 (18.92%) |  |
| 2 - 5 | 8 (21.62%) | 6 (28.57%) |  | 39 (69.64%) | 30 (81.08%) |  |
| **Tumor location** |  |  | 0.621 |  |  | 0.671 |
| Right lobe | 30 (81.08%) | 16 (76.19%) |  | 40 (71.43%) | 27 (72.97%) |  |
| Left lobe | 5 (13.51%) | 4 (19.05%) |  | 14 (25.00%) | 8 (21.62%) |  |
| Bilobar | 2 (5.41%) | 1 (4.76%) |  | 1 (1.77%) | 1 (2.71%) |  |
| **Perivascular location** |  |  | 0.853 |  |  | 0.885 |
| Yes | 5 (13.51%) | 2 (9.52%) |  | 11 (19.64%) | 7 (18.92%) |  |
| No | 32 (86.49%) | 19 (90.48%) |  | 45 (80.36%) | 30 (81.08%) |  |
| **Located in periphery of liver** |  |  | 0.796 |  |  | 0.671 |
| Yes | 7 (18.92%) | 4 (19.05%) |  | 24 (42.86%) | 18 (48.65%) |  |
| No | 30 (81.08%) | 17 (80.95%) |  | 32 (57.14%) | 19 (51.35%) |  |
| **ALBI scores** |  |  | 0.438 |  |  | 0.614 |
| 1 | 18 (48.65%) | 10 (47.62%) |  | 20 (35.71%) | 19 (51.35%) |  |
| 2 | 18 (48.65%) | 9 (42.86%) |  | 35 (62.50%) | 18 (48.65%) |  |
| 3 | 1 (2.70%) | 2 (9.52%) |  | 1 (1.79%) | 0 (0.00%) |  |

RFA, radiofrequency ablation; SR, surgical resection; 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.

**Table S3.** The structure and hyper-parameters of Depthwise convolution.

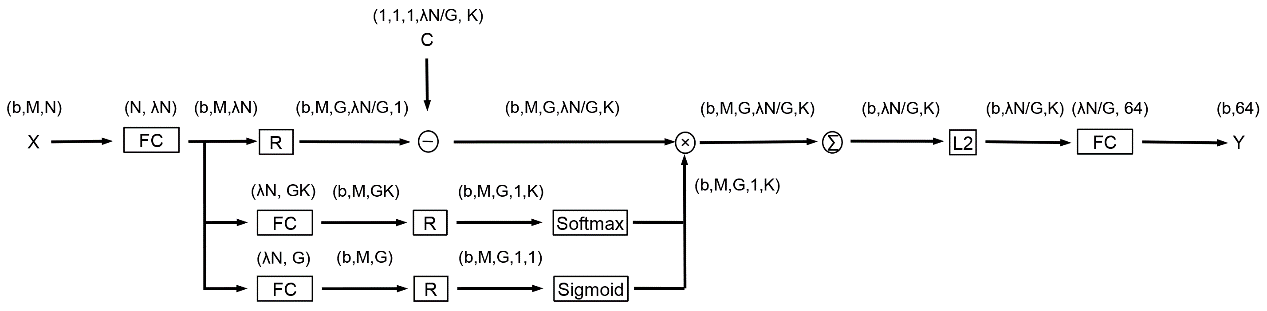
|  |  |  |  |
| --- | --- | --- | --- |
| **Layers** | **Input size** | **Output size** | **Hyper-parameters** |
| **Convolution 1** | Ci×H×W | Ci×H×W | Kernel: K×K Stride: 1×1  Padding: |K/2|×|K/2| Groups: Ci |
| **Convolution 2** | Ci×H×W | Co×H×W | Kernel: 1×1 Stride: 1×1  Padding: 1×1 Groups: 1 |

Ci: the channel size of input data; Co: the channel size of output data; H: the height of data; W: the width of data; K: the size of convolution kernel, which is the main hyper-parameter of depthwise convolution module.

**Table S4.** The data size and hyper-parameters of each layer.

|  |  |  |  |
| --- | --- | --- | --- |
| **Layers** | **Input size** | **Output size** | **Hyper-parameters** |
| **Convolution** | AP: 44×1×64×64  PP: 80×1×64×64  VP: 36×1×64×64 | AP: 44×16×32×32  PP: 80×16×32×32  VP: 36×16×32×32 | Kernel: 5×5 Stride: 2×2  Padding: 2×2 Groups: 1  Ci: 1 Co: 16 |
| **ReLU** | AP: 44×16×32×32  PP: 80×16×32×32  VP: 36×16×32×32 | AP: 44×16×32×32  PP: 80×16×32×32  VP: 36×16×32×32 | No hyper-parameter |
| **Convolution** | AP: 44×16×32×32  PP: 80×16×32×32  VP: 36×16×32×32 | AP: 44×16×32×32  PP: 80×16×32×32  VP: 36×16×32×32 | Kernel: 3×3 Stride: 1×1  Padding: 1×1 Groups: 1  Ci: 16 Co: 16 |
| **ReLU** | AP: 44×16×32×32  PP: 80×16×32×32  VP: 36×16×32×32 | AP: 44×16×32×32  PP: 80×16×32×32  VP: 36×16×32×32 | No hyper-parameter |
| **MaxPooling** | AP: 44×16×32×32  PP: 80×16×32×32  VP: 36×16×32×32 | AP: 44×16×16×16  PP: 80×16×16×16  VP: 36×16×16×16 | Kernel: 2×2 Stride: 2×2 |
| **Depthwise convolution** | AP: 44×16×16×16  PP: 80×16×16×16  VP: 36×16×16×16 | AP: 44×32×16×16  PP: 80×32×16×16  VP: 36×32×16×16 | Kernel: 3×3  Ci: 16 Co: 32 |
| **ReLU** | AP: 44×32×16×16  PP: 80×32×16×16  VP: 36×32×16×16 | AP: 44×32×16×16  PP: 80×32×16×16  VP: 36×32×16×16 | No hyper-parameter |
| **MaxPooling** | AP: 44×32×16×16  PP: 80×32×16×16  VP: 36×32×16×16 | AP: 44×32×8×8  PP: 80×32×8×8  VP: 36×32×8×8 | Kernel: 2×2 Stride: 2×2 |
| **Depthwise convolution** | AP: 44×32×8×8  PP: 80×32×8×8  VP: 36×32×8×8 | AP: 44×64×8×8  PP: 80×64×8×8  VP: 36×64×8×8 | Kernel: 3×3  Ci: 32 Co: 64 |
| **ReLU** | AP: 44×64×8×8  PP: 80×64×8×8  VP: 36×64×8×8 | AP: 44×64×8×8  PP: 80×64×8×8  VP: 36×64×8×8 | No hyper-parameter |
| **MaxPooling** | AP: 44×64×8×8  PP: 80×64×8×8  VP: 36×64×8×8 | AP: 44×64×4×4  PP: 80×64×4×4  VP: 36×64×4×4 | Kernel: 2×2 Stride: 2×2 |
| **Depthwise convolution** | AP: 44×64×4×4  PP: 80×64×4×4  VP: 36×64×4×4 | AP: 44×96×4×4  PP: 80×96×4×4  VP: 36×96×4×4 | Kernel: 3×3  Ci: 64 Co: 96 |
| **ReLU** | AP: 44×96×4×4  PP: 80×96×4×4  VP: 36×96×4×4 | AP: 44×96×4×4  PP: 80×96×4×4  VP: 36×96×4×4 | No hyper-parameter |
| **MaxPooling** | AP: 44×96×4×4  PP: 80×96×4×4  VP: 36×96×4×4 | AP: 44×96×2×2  PP: 80×96×2×2  VP: 36×96×2×2 | Kernel: 2×2 Stride: 2×2 |
| **Global AvgPooling** | AP: 44×96×2×2  PP: 80×96×2×2  VP: 36×96×2×2 | AP: 44×96×1×1  PP: 80×96×1×1  VP: 36×96×1×1 | Kernel: 2×2 Stride: 2×2 |
| **Concatenate** | AP: 44×96×1×1  PP: 80×96×1×1  VP: 36×96×1×1 | 160×96 | No hyper-parameter |
| **VLAD** | 160×96 | 64 | See Figure S1 for details |
| **Cox Regression** | 64 | 1 | Ci: 64 Co: 1 |

Ci: the channel size of input data; Co: the channel size of output data; AP, arterial phase; PP, portal phase; VP, venous phase.



**Fig. S1.** The design of global aggregation VLAD module. The frame-level feature X is encoded to one common feature space by FC, and then mapped to 2 feature maps for computing weight maps in inter-groups (Softmax branch) and intra-groups (Sigmoid branch) to strengthen the weights of useful features and weaken the weights of useless features. The difference between encoded X and learned cluster matrix C was compared and aggregated by computed weights maps. Finally the resulting features was normalized and mapped to a low-dimension feature Y. We annotated the hyper-parameters of all layers with learning parameters above boxes. The size of data in each processing phase was annotated above the arrow. We relied on Pytorch’s broadcasting semantics if there was difference between the size of input data during addition, subtraction and multiplication. The Softmax and L2 functions were applied on the last dimension of data. The hyper-parameters of VLAD and the corresponding values are: b: 128 when training and 1 when validating; M: 160; N: 96; λ: 2; G: 2; K: 8. FC, fully connection; R, reshape; L2, normalization by restricting the norm to be 1; X, the input of all frames’ features; Y, the output of global aggregated feature.

**Supplementary Figure 2**

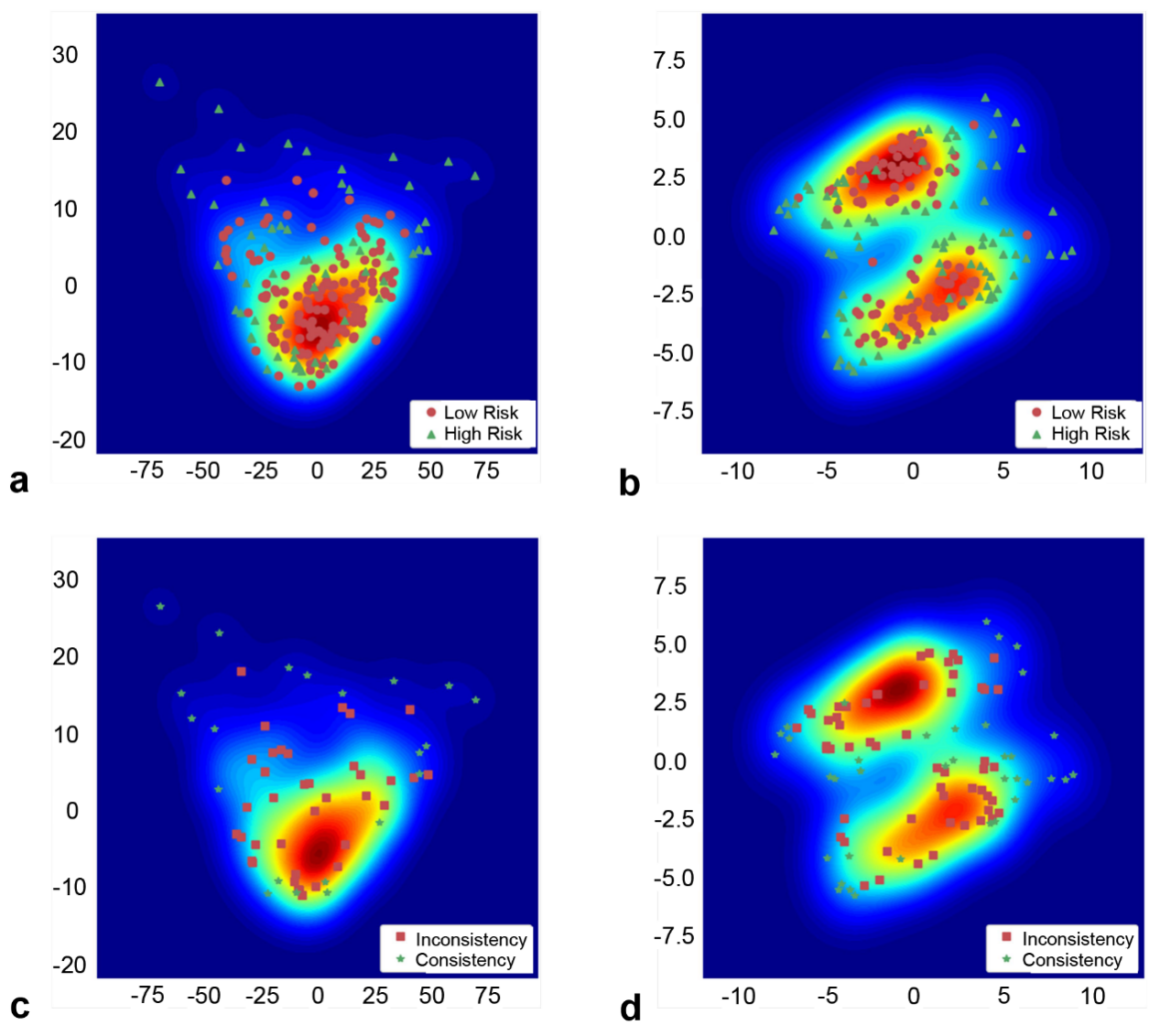
**Fig. S2.** Visualization of Radiomics features from R-RFA and R-SR. (a) The feature visualization using R-RFA in AP, PP and VP. (b) The same visualization using R-SR. Bounding box ROIs of CEUS cines in all frames were converted into pseudo-colored feature maps for R-RFA and R-SR, respectively. Areas with warm colors (e.g., red and yellow) represent stronger correlation with the prognostic prediction. In contrast, cold color areas (e.g., green and blue) indicate weaker correlation for the prediction. ROI, region of interest; CEUS, contrast-enhanced ultrasound; AP, arterial phase; PP, portal phase; VP, venous phase.

Supplementary Figure 3

**Fig. S3.** Kaplan-Meier survival curves of individualized models. (a)(b) represent the Kaplan-Meier survival curves of the individualized model of RFA in the training and validation cohorts, and (c)(d) for the individualized model of SR. The short vertical lines indicate censored data. ROC, receiver operator characteristic; PFS, progression-free survival; N-RFA, the nomogram for RFA; N-SR, the nomogram for SR.

Supplementary Figure 4

**Fig. S4.** Time-dependent ROC curves.(a)(b) ROC curves given by the R-RFA and R-SR model for two-year PFS in their corresponding training and validation cohorts, respectively. (c)(d) ROC curves given by the individualized models of RFA and SR for two-year PFS in their corresponding training and validation cohorts, respectively. ROC, receiver operator characteristic; PFS, progression-free survival; N-RFA, the nomogram for RFA; N-SR, the nomogram for SR.



**Fig. S5.** The relations between network features density distribution and the distribution of different kinds of patient subgroups. The distribution of low-/high-risk patients in the network features space for patients in RFA group (a) and SR group (b). The distribution of inconsistency/consistency patients in the network features space for patients in RFA group (c) and SR group (d).