**Supplementary Data**

**S1. Flow chart of the study population**



Note. HCC, hepatocellular carcinoma; MRI, magnetic resonance imaging; TACE, transcatheter arterial chemoembolization; RFA, radiofrequency ablation; CTX, chemotherapy; CRT, chemoradiotherapy; HBP, hepatobiliary phase; SI, signal intensity

**S2. Magnetic resonance image Acquisition parameters**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Acquisition Sequence** | **Scanner Trade Name (Field Strength)** | **Matrix Size** | **Section Thickness (mm)** | **Intersection Gap (mm)** | **Repetition Time (msec)** | **Echo Time (msec)** | **Flip Angle (°)** |
| Dual–echo T1–weighted gradient–recalled-echo | Magnetom Trio Tim (3T) | 256ⅹ192 | 3 | 0.6 | 4 | 2.46 and 1.23 | 9 |
| Intera Achieva (3T) | 160ⅹ160 | 5 | 1 | 219 | 2.3 and 1.1 | 20 |
| Intera Achieva (1.5T) | 288ⅹ231 | 7 | 0.7 | 189 | 4.6 and 2.3 | 80 |
| Discovery MR750w (3T) | 320ⅹ250 | 4 | 1 | 4.9 | 2.3 and 1.1 | 10 |
| T1–weighted 3D gradient–echo with dynamic contrast enhancement† | Magnetom Trio Tim (3T) | 256ⅹ192 | 2 | 0.4 | 2.54 | 0.95 | 13 |
| Intera Achieva (3T) | 256ⅹ220 | 2 | 0 | 3 | 1.42 | 10 |
| Intera Achieva (1.5T) | 256ⅹ166 | 4 | 0 | 4.2 | 2.0 | 10 |
| Discovery MR750w (3T) | 224ⅹ192 | 2.5 | 0 | 190 | 1.2–4.7 | 10 |
| T2–weighted turbo spin-echo, navigator–triggered | Magnetom Trio Tim (3T) | 320ⅹ168 | 2 | 0.4 | 2.54 | 0.95 | 140 |
| Intera Achieva (3T) | 288ⅹ192 | 5 | 1 | 988 | 80 | 90 |
| Intera Achieva (1.5T) | 288ⅹ228 | 7 | 0 | 452 | 80 | 90 |
| Discovery MR750w (3T) | 320ⅹ320 | 4 | 1 | 4286 | 73 | 15 |
| Diffusion–weighted imaging | Magnetom Trio Tim (3T) | 128ⅹ96 | 5 | 1 | 5200 | 67 | 90 |
| Intera Achieva (3T) | 288ⅹ192 | 5 | 1 | 8500 | 57 | 90 |
| Intera Achieva (1.5T) | 112ⅹ144 | 7 | 1 | 1600 | 62 | 90 |
| Discovery MR750w (3T) | 96ⅹ128 | 4 | 1 | 5800 | 62 | 90 |

† The arterial phase began 2 or 3 seconds after the peak aortic enhancement was determined with the bolus injection; subsequent dynamic images were obtained at intervals of approximately 30 seconds. Each dynamic image acquisition required 18–24 seconds. Hepatobiliary images were obtained 15 or 20 minutes after injection of the contrast material by using the same imaging sequence as that used for the pre- and postcontrast images.

**S3. Three prognostic models for predicting disease-free survival in hepatocellular carcinoma**

Three-dimensional segmentation was performed in the hepatobiliary phase (HBP) image of gadoxetic acid-enhanced MRI. Because of the high contrast difference in HBP, the objectivity of segmentation of the tumor increased. The radiomic prognostic model was built using the extracted 3D radiomic features, and the clinicopathologic model was built using the independent prognostic factors among the clinical and surgical pathologic information. The combined clinicopathologic-radiomic model was built by incorporating the radiomic model into the clinicopathologic model.



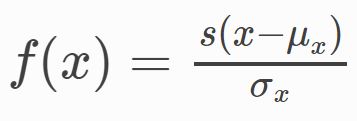
**S4. Detailed methods of radiomic analysis**

***Data acquisition, image segmentation, and preprocessing***

In the preoperative gadoxetic acid contrast-enhanced dynamic liver, MRI, AP, PP, and HBP images were downloaded in a Digital Imaging and Communications in Medicine format, and registration was performed on the three phases using 3D slicer (www.slicer.org). Voxels in each MR image volume were resampled to an isotropic voxel size of 1.0 × 1.0 × 1.0 mm3 to minimize the effects of different MR imaging conditions, such as pixel spacings and slice thicknesses. Semi-automatic 3-dimensional segmentation of tumor was performed in HBP images using FastGrowCut effect of 3D slicer, and the generated mask was commonly used for AP, PP, and HBP images. The 3D mask of tumor drawn on the HBP-hypointense area expanded the 3 mm and 5 mm radially outwards using the dilate effect of the 3D slicer. Consequently, three kinds of 3D masks were created according to the extended length of the tumor (0 mm, 3 mm, and 5 mm). This process is called tumor border extension and is essential for extracting the radiomic features of the peritumoral area.

***Radiomic feature extraction***

MRI signal intensity normalization and feature extraction were performed using PyRadiomics (Computational Imaging and Bioinformatics Lab, Harvard Medical School). PyRadiomics enabled the normalization of image intensity values prior to feeding them to the feature extraction pipeline. Normalization centered the image at the mean with standard deviation. Normalization is based on the all gray values contained within the image, not just those defined by the volume of interest (VOI) in the mask.



Where:

* x and f(x) are the original and normalized intensity, respectively.
* μx and σx are the mean and standard deviation of the image intensity values.
* s is a scaling factor. s was set to 100 in this study.

In addition, outliers were removed, in which case values for which x > μx + 3σx or x < μx - 3σx were set to μx + 3σx and μx - 3σx, respectively. Removal of outliers was done after the values of the image were normalized, but before scale was applied. Then, all voxel values were shifted by 300 to ensure that all voxels had positive values. For gray value discretization, bin width was set to 5, which means the number of bins was set to 600 / 5 = 125. All the processes about normalization and gray value discretization were performed using PyRadiomics.

Extracted feature classes consisted of the first-order features, shape features, the second-order features, and two high order filters — Laplacian of Gaussian (LoG) and Wavelet filter.

First-order statistics described the distribution of voxel intensities within the image region defined by the mask through commonly used and basic metrics. The second-order features consisted of Gray Level Co-occurrence Matrix (GLCM) Features, Gray Level Size Zone Matrix (GLSZM) Features, Gray Level Run Length Matrix (GLRLM) Features, Neighboring Gray Tone Difference Matrix (NGTDM) Features, and Gray Level Dependence Matrix (GLDM) Features.

Applied filter classes were Laplacian of Gaussian (LoG) and Wavelet filter. LoG filter is an edge enhancement filter, which emphasizes areas of gray level change, where sigma defines how coarse the emphasized texture should be. A low sigma emphasis on fine textures (change over a short distance), where a high sigma value emphasizes coarse textures (gray level change over a large distance). Five sigma values were used in this study (1 mm, 2 mm, 3 mm, 4 mm, and 5 mm). Wavelet filtering yields 8 decompositions per level (all possible combinations of applying either a high or a low pass filter in each of the three dimensions).

The following were extracted from the lesion: 13 shape features, 18 first-order features, 74 second-order features with LoG filters with 5 sigmas, and 8 combinations of wavelet filter. This led to 1301 features for one phase and 3903 features for all three dynamic phases in total. Radiomic features were extracted for each combination of AP, PP, and HBP phases and for each border extension of 0 mm, 3 mm, and 5 mm, respectively.

***Radiomic feature normalization and feature clustering***

All features were normalized on the training cohort using z-score normalization. The resulting scale and shift constants used in the z-score normalization process were saved and then applied to the feature normalization of the independent validation cohort. After normalization, hierarchical feature clustering was performed on the training cohort using Spearman correlation coefficient (SCC) to remove redundant radiomic features. Highly correlated imaging biomarkers (SCC > 0.80) were clustered using hierarchical clustering. The resulting cluster was represented by a representative feature, which had the largest range of values among the clustered features. Since the radiomic feature was normalized using z-score after feature extraction, the range of feature means the range of normalized values of feature. The cluster configurations and representative features induced from the training cohort were applied to the validation cohort. Because all the parameters, such as z-normalization parameters, clustering configuration, and representative features, were part of the radiomic model, they must be remembered with the model building and applied to the validation set with the same cutoff. All processes related to radiomic feature normalization and clustering were performed using R (version 3.3.1; R Foundation for Statistical Computing, Vienna, Austria).

***Radiomic feature selection***

Radiomic feature selection was performed using a random forest minimal depth algorithm, one of the feature extraction methods of dataset for survival analysis. The top 20 best ranking features were selected and used for further analysis. All processes related to radiomic feature selection were performed using R (version 3.3.1; R Foundation for Statistical Computing, Vienna, Austria) and R package (RandomForestSRC).

***Hyperparameter optimization***

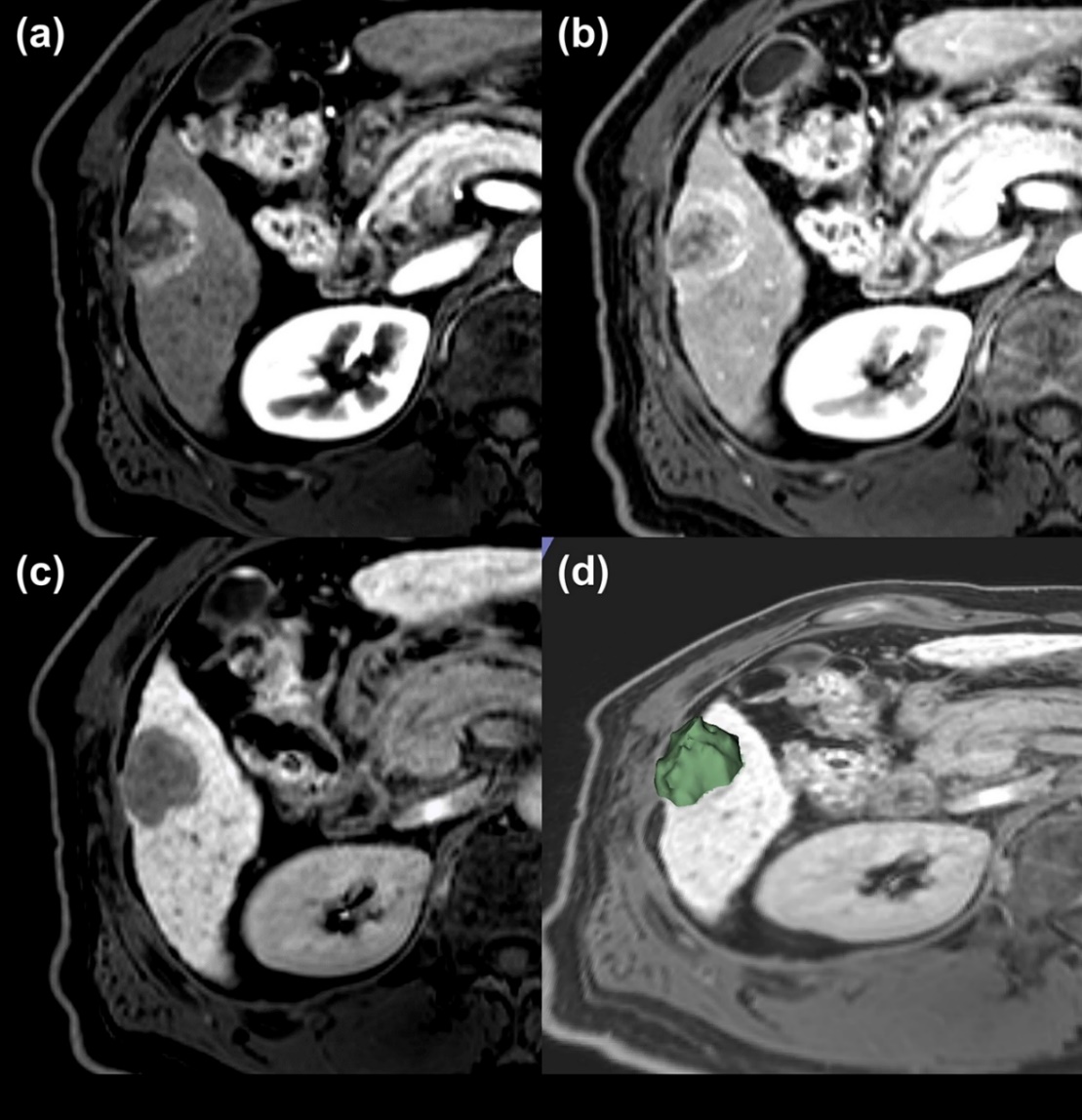
Hyperparameter optimization was performed to increase model generalizability before building the final prognostic model. Hyperparameters of the random survival forest, such as the number of features that make radiomic signature or algorithm-specific settings, were optimized. A major objective of hyperparameter tuning is to limit model overfitting and obtain generalized model. Overfitting would otherwise lead to poor predictive performance on unseen data. The hyperparameters were tuned using nested cross-validation with outer 5-fold and inner 2-fold scheme based on the training cohort. Hyperparameter optimization was performed using grid search through a predefined hyperparameter space. The purpose of hyperparameter optimization was to find a combination of hyperparameters with the best aggregated performance of the internal validation fold. Importantly, any data used in the hyperparameter optimization should not overlap the validation cohort to maintain the independency of the validation cohort. All processes related to hyperparameter optimization were performed using R (version 3.3.1; R Foundation for Statistical Computing, Vienna, Austria) and R package (mlr, ranger).

***Radiomic model building and validation***

Model training was performed n = 1000 times using bootstrap samples of the training cohort. Two algorithms based on random survival forest—the random survival forest with logrank splitting (logrank) method and the random survival forest with the maximally selected rank statistics (maxstat) method, which have recently reported good results as model training methods for survival analysis—were used as model learning. These algorithms were learned using selected features and hyperparameters. We presented both results of two different training algorithms to show the generality of the results.

The model performance in training cohort and validation cohort was assessed using the concordance index (c-index). The c-index value for the training set was the out-of-bag c-index value in the random survival forest. However, the performance of the training cohort is generally not used as model performance because of the risk of overfitting. Thus, all the model performance presented in this study used the performance result from the independent validation set. The final model performances were reported by averaging the performances in the independent validation cohort using the models generated by bootstrap resampling (n = 1000). All processes related to the model building and validation were performed using R (version 3.3.1; R Foundation for Statistical Computing, Vienna, Austria) and R package (mlr, ranger, boot).

**S5. Case Figure.** Magnetic resonance imaging findings for a representative case involving a 54-year-old male exhibiting a 3.0-cm hepatocellular carcinoma (HCC). Gadoxetic acid-enhanced T1-weighted MR images were obtained during the (a) arterial phase, (b) portal venous phase, and (c) hepatobiliary phase (HBP). Three-dimensional segmentation of MR HBP images of HCC was performed using 3D-slicer (d). The HCC was pathologically accessed as negative microvascular invasion, positive capsule formation with underlying hepatitis B-related cirrhosis. The HCC was classified as a low-risk group in the clinicopathologic prognostic model, but it was classified as a high-risk group in the radiomic model and combined clinicopathologic-radiomic model for early recurrence. The HCC showed poor prognosis in that it recurred in 20 months after surgery.

****

**S6. Selected radiomic features and Reproducibility**

***Selected features for 2-5 cm HCCs with 3 mm border extension***

The prefix *meta-* means that the feature is the representative feature selected in the clustered feature group. The feature clustering group was created by hierarchical clustering based on Spearman's correlation coefficient. The representative feature had the largest range of values among the clustered features. Since the radiomic feature was normalized using z-score after feature extraction, the range of feature means the range of normalized values of feature. In the case of a representative feature, the list of features was described below the representative feature. The order of importance of the selected features was plotted according to the minimum depth order of the random survival forest.

|  |  |  |
| --- | --- | --- |
| **Selected Features** | **ICC** | **ICC CI** |
| [1] wavelet.HHL\_glszm\_GrayLevelNonUniformityNormalized | 0.988 | 0.975, 0.994 |
| [2] meta\_wavelet.HHH\_glszm\_ZoneVariance | 0.813 | 0.646, 0.906 |
| wavelet.HHH\_glszm\_SmallAreaLowGrayLevelEmphasis  wavelet.HHH\_glszm\_ZoneVariance |  |  |
| [3] meta\_log.sigma.1.0.mm.3D\_glcm\_Correlation | 0.996 | 0.992, 0.998 |
| log.sigma.1.0.mm.3D\_glcm\_Correlation  log.sigma.1.0.mm.3D\_glcm\_Imc2 |  |  |
| [4] meta\_log.sigma.5.0.mm.3D\_glszm\_SizeZoneNonUniformityNormalized.1 | 0.825 | 0.667, 0.912 |
| log.sigma.5.0.mm.3D\_glszm\_SizeZoneNonUniformityNormalized.1  log.sigma.5.0.mm.3D\_glszm\_SmallAreaEmphasis.1 |  |  |
| [5] meta\_log.sigma.2.0.mm.3D\_glcm\_Imc2.1 | 0.994 | 0.988, 0.997 |
| log.sigma.2.0.mm.3D\_glcm\_Correlation.1  log.sigma.2.0.mm.3D\_glcm\_Imc2.1 |  |  |
| [6] log.sigma.1.0.mm.3D\_glcm\_InverseVariance | 0.999 | 0.998, 0.999 |
| [7] wavelet.HHL\_glszm\_SizeZoneNonUniformityNormalized.1 | 0.911 | 0.822, 0.956 |
| [8] meta\_log.sigma.4.0.mm.3D\_glcm\_Correlation.2 | 0.997 | 0.995, 0.999 |
| log.sigma.4.0.mm.3D\_glcm\_Correlation.2  log.sigma.4.0.mm.3D\_glcm\_Imc2.2  log.sigma.5.0.mm.3D\_glcm\_Correlation.2  log.sigma.5.0.mm.3D\_glcm\_Imc2.2 |  |  |

Note. ICC, intraclass correlation coefficient; CI, confidence interval

