**Supplementary Methods**

1. **The diagnostic criteria of cirrhosis and hepatocellular carcinoma (HCC)**

***1.1 HCC***

The diagnosis of HCC should meet any 1 of the following criteria:

1) Histopathological confirmation by liver biopsy or surgically excised tissue;

2) Detection of a lesion which meets criteria for HCC with at least two imaging techniques (trans-abdominal ultrasound (US), triphasic abdominal computed tomography (CT) scan, magnetic resonance imaging study of liver, or hepatic angiogram);

3) Detection with one imaging technique coupled with an alpha-fetoprotein concentration greater than 400 μg/L.

***1.2 Cirrhosis***

The criteria for hepatic cirrhosis are defined as follows (including compensated and decompensated):

1) Evidence of cirrhosis based on liver biopsy (evaluated by Ishak score system);

2) In the absence of liver biopsy results, the patient should meet 1 of the following criteria:

− Ascites;

− Hepatic encephalopathy;

− Upper gastro-intestinal (esophageal and/or gastroduodenal) bleeding;

− Hepatorenal syndrome;

The complication caused by non-cirrhotic portal hypertension (such as portal vein thrombosis, congenital hepatic fibrosis, idiopathic portal hypertension, etc.) should be excluded;

3) If none of the criteria above are met, the patient should meet any 2 of the following 4 criteria:

− Liver imaging showing features of cirrhosis: nodular liver and/or splenomegaly;

− Platelet count <150,000/mm3 in the absence of other explanation;

− Liver stiffness measurement (by Fibroscan) >12 kPa;

− Gastro-esophageal varices as visualized by upper endoscopy.

1. **Data and modeling**

***2.1 Data***

To ensure better generalization of the model, patients from multiple centers with evident or underlying variances were divided into the training and validation cohorts at a ratio of 7:3, while patients from one center with the largest sample size (Nanfang Hospital) were assigned as the test cohort. The cut-off for data to be included in the analysis was January 2024.

***CT image acquisition and pre-processing***

All patients included underwent 3-phase contrast-enhanced abdominal CT at enrollment. Liver and spleen CT images from the arterial, venous, and delay phases were retrieved from the Picture Archiving and Communication System (PACS) and subjected to analysis. Variations in CT systems, scan parameters (including detector collimation, post-injection time for the three phases), and image characteristics across different centers were detailed in Supplementary Table 2.

Raw images acquired in DICOM format were converted to NIfTI format using dcm2niix (<https://github.com/rordenlab/dcm2niix>)1. The relatively heterogeneous resolution in the z-axis, due to inconsistent detector collimation and reconstruction, does not permit a meaningful and reliable 3D analysis of the images. Therefore, the images were resampled to a consistent spatial resolution of 1.0 × 1.0 × 5.0 mm by PyRadiomics package in Python, version 3.9 (https://www.python.org)2. Subsequently, we adjusted the Hounsfield units (HU) to a window of [-170, 230] to enhance the contrast of the liver and spleen, and normalized the HU values.

***Clinical data collection and aMAP score calculation***

Laboratory test results were recorded at each follow-up visit. The aMAP score, a regression-based predictive model, was calculated for each patient at every visit as follows:

$$aMAP score = ((0.06 × age + 0.89 × sex (male: 1, female: 0) + 0.48 × ((log\_{10}total bilirubin × 0.66) + (albumin ×-0.085)) -0.01 × platelets) +7.4)/14.77 × 100$$

, where age is in years, total bilirubin in μmol/L, platelets in 103/mm3, and albumin in g/L3.

***2.2 Segmentation***

To comprehensively capture and independently validate the characteristics of the liver and spleen, we defined each entire organ as a separate ROI. We employed an automatic segmentation method known as nnU-Net, which analyzes the provided training cases and automatically configures a matching U-Net-based segmentation pipeline4.

Considering the large size of our dataset, fully manual mask segmentation was not feasible. To address this, we employed a two-step segmentation process: pre-training and formal training.

For pre-training, we used the task03\_Liver (n=201) and task09\_Spleen (n=61) datasets from the Medical Segmentation Decathlon (<http://medicaldecathlon.com>) as the training set, while images from 300 patients at our center served as the test set. This step produced initial liver and spleen masks for these 300 cases. Radiologists from Nanfang Hospital then reviewed and manually refined these masks using ITK-SNAP (version 4.0.1, [www.itksnap.org](http://www.itksnap.org)) to ensure precise alignment with the original organ boundaries. The refined masks and their corresponding images were subsequently used as the training set for the formal training phase. The remaining cases were designated as the testing set, and the masks for these cases were then predicted. To assess the effectiveness of the segmentation process in the formal training phase, we evaluated performance using the Dice score.

All images and masks mentioned above are 3-dimentional. The U-Net configuration was 3d\_fullres and the model was trained on 3-fold cross-validation.

***2.3 Model construction***

***Feature extraction***

Radiomics features were extracted from the 3-stage 3D ROIs using the PyRadiomics package in Python, version 3.92. Details regarding radiomics feature extraction were followed The Image Biomarker Standardization Initiative (IBSI) guidelines (Supplementary Table 3)5.

Deep learning features consisted of the last fully connected (FC) layer of a fine-tuned ResNet-18 model pretrained on the ImageNet dataset (https://image-net.org). Comprehensive explanations of the deep learning feature extraction processes are provided below:

**Network architecture：**We proposed fine-tuned ResNet-18 as the deep learning model. It is universally recognized that with the network depth increasing, accuracy gets saturated and then degrades rapidly. Kaiming He team addressed the degradation problem by introducing a deep residual learning framework called ResNet, which won first places on the tasks of ImageNet detection, ImageNet localization, COCO detection, and COCO segmentation6. After preprocessing, to ensure the consistency of deep learning network input, every slice in NIfTI files was resized to $32×128×128$ pixels. During model training, since our expected output binary classification label (i.e., tumorigenesis and non-tumorigenesis), the dimension of the last full-connected (FC) layer was altered to $512×2$. The weight file with best accuracy in validation cohort was saved as the best model. Then we load the weights file of the best model to run all cases in three cohorts, when the full-connected layer (the original last layer of network structure) was removed, and the weights of average-pool layer (the second last layer) were extracted as features. Every case has 512 features extracted. The Network architecture was shown in Supplementary Fig. 1. Deep models were trained exclusively on training cohort, and validation cohort was used to adjust and select the best hyperparameters.

**Implementation details:** The loss function of the HCC occurrence prediction is binary cross-entropy. To minimize the loss function, we use Adam algorithm to obtain the optimal parameters. The learning rate is set at 10-2 initially. The pretrained ResNet-18 model was trained for 400 epochs with a batch size of 512. We trained the model using Python on NVIDIA GeForce RTX 4090, 13th Gen Intel® Core™ i9-13900K, 3.00 GHz and 128 GB of internal memory.

***Feature selection and image signature score construction***

We applied LASSO (Least Absolute Shrinkage and Selection Operator) regression for liver and spleen image features selection. These selected features were used to construct the image score via logistic regression. ResNet-18, LASSO, and logistic models were all trained exclusively on training cohort, and validation cohort was used to adjust and select the best hyperparameters. Deep features and predictive scores were generated subsequently for all three cohorts using models with the best parameters.

***Model integration***

The image signature score was added to the aMAP model to enhance model performance using logistic regression. Due to the imbalance between HCC and non-HCC patients in our dataset, we employed the 'class\_weight="balanced"' parameter when performing logistic regression. The combined model represented the final model, thereafter called the “aMAP-CT model”.

***2.4 Model evaluation***

To evaluate the classification performance and predictive ability of the aMAP-CT model, we analyzed the area under the receiver operating characteristic curve (AUC) and used Net Reclassification Improvement (NRI) to assess reclassification accuracy compared to the aMAP model. Calibration plots illustrated the agreement between predicted and actual HCC probabilities, while sensitivity, specificity, accuracy, F1-score, Negative Predictive Value (NPV), and Positive Predictive Value (PPV) were calculated.

Two risk groups of “high” and “low” at the cut-off point were identified by surv\_cutpoint function in survminer package from R software. We assessed the model's prognostic value by calculating the 3-year cumulative incidence of HCC using the Kaplan-Meier (K-M) method, with comparisons made using the log-rank test. The standardized ‘net benefit’ was estimated from decision curve analysis (DCA). Subgroup analyses evaluated predictive capability across different subsets. We also compared the AUC of aMAP-CT with existing HCC risk scores to demonstrate its superiority. A Sankey plot was generated online (<https://dycharts.com/appv2/#/pages/home/index>).

**Supplementary References**

[1] Li X, Morgan PS, Ashburner J, Smith J, Rorden C. The first step for neuroimaging data analysis: DICOM to NIfTI conversion. J Neurosci Methods 2016;264:47-56. doi: 10.1016/j.jneumeth.2016.03.001. PMID: 26945974

[2] van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic Phenotype. Cancer Res 2017;77(21):e104-e107. doi: 10.1158/0008-5472.Can-17-0339. PMID: 29092951

[3] Fan R, Papatheodoridis G, Sun J, Innes H, Toyoda H, Xie Q, et al. aMAP risk score predicts hepatocellular carcinoma development in patients with chronic hepatitis. Journal of hepatology 2020;73(6):1368-1378. doi: 10.1016/j.jhep.2020.07.025. PMID: 32707225

[4] Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods 2021;18(2):203-211. doi: 10.1038/s41592-020-01008-z. PMID: 33288961

[5] Zwanenburg A, Vallières M, Abdalah MA, Aerts H, Andrearczyk V, Apte A, et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. Radiology 2020;295(2):328-338. doi: 10.1148/radiol.2020191145. PMID: 32154773

[6] He K, Zhang X, Ren S, Sun JJae-p. Deep Residual Learning for Image Recognition. 2015. <https://ui.adsabs.harvard.edu/abs/2015arXiv151203385H> (accessed December 01, 2015).