CT images-based 3D convolutional neural network to predict early recurrence of solitary hepatocellular carcinoma after radical hepatectomy

Hao Cui*
Kun-Yuan Wang*
Wen-Yuan Li*
Hong-Bo Zhu
Lu-Shan Xiao*
Li Liu*

PURPOSE
The high rate of recurrence of hepatocellular carcinoma (HCC) after radical hepatectomy is an important factor that affects the long-term survival of patients. This study aimed to develop a computed tomography (CT) images-based 3-dimensional (3D) convolutional neural network (CNN) for the preoperative prediction of early recurrence (ER) (≤2 years) after radical hepatectomy in patients with solitary HCC and to compare the effects of segmentation sampling (SS) and non-segmentation sampling (NSS) on the prediction performance of 3D-CNN.

METHODS
Contrast-enhanced CT images of 220 HCC patients were used in this study (training group = 178 and test group = 42). We used SS and NSS to select the volume-of-interest to train SS-3D-CNN and NSS-3D-CNN separately. The prediction accuracy was evaluated using the test group. Finally, gradient-weighted class activation mappings (Grad-CAMs) were plotted to analyze the difference of prediction logic between the SS-3D-CNN and NSS-3D-CNN.

RESULTS
The areas under the receiver operating characteristic curves (AUCs) of the SS-3D-CNN and NSS-3D-CNN in the training group were 0.824 (95% CI: 0.764-0.885) and 0.868 (95% CI: 0.815-0.921). The AUC of the SS-3D-CNN and NSS-3D-CNN in the test group were 0.789 (95% CI: 0.637-0.941) and 0.560 (95% CI: 0.378-0.742). The SS-3D-CNN could stratify patients into low- and high-risk groups, with significant differences in recurrence-free survival (RFS) ($P < .001$). But NSS-3D-CNN could not effectively stratify them in the test group. According to the Grad-CAMs, compared with SS-3D-CNN, NSS-3D-CNN was obviously interfered by the nearby tissues.

CONCLUSION
SS-3D-CNN may be of clinical use for identifying high-risk patients and formulating individualized treatment and follow-up strategies. SS is better than NSS in improving the performance of 3D-CNN in our study.

From the Department of Medical Quality Management (H.C., L.L., ✉ liuli@i.smu.edu.cn); Hepatology Unit and Department of Infectious Diseases (H.C., K.-Y.W., L.-S.X ✉ 15622178423@163.com, L.L.); China Big Data Center (W.-Y.L.), Nanfang Hospital, Southern Medical University, Guangzhou, China; and Department of Oncology (H.-B.Z.), The First Affiliated Hospital, University of South China, Hengyang, China.

*Hao Cui, Kun-Yuan Wang, Lu-Shan Xiao, and Li Liu contributed equally to this work.

Hepatocellular carcinoma (HCC) is the fourth most common cause of cancer-related death worldwide and accounts for >80% of primary liver cancers, which constitutes a major global health problem.[1,2] At present, there are many available treatment methods for HCC. Hepatic resection is still the treatment of choice for early HCC in patients with well-preserved liver function.[3] However, the rate of recurrence after radical hepatectomy is high.[3-5] Five-year recurrence rates approach 70% post-resection, with two-third of recurrences related to intrahepatic spread occurring within two years.[6] A cutoff of two years has been adopted to grossly classify early recurrence (ER) and late recurrence.[7] Consequently, if clinicians can identify high-risk ER groups before resection, they can consider liver transplant or adjuvant treatment after resection to reduce the recurrence rate and develop closer follow-up strategies.[8]

In previous studies, it has been confirmed that computed tomography (CT), magnetic resonance (MR), positron emission tomography (PET), and other medical images are data more than images,[9,10] which can be used to develop models.[11,12] Since recent years, deep learning algorithms have been widely used in medical imaging. In particular, the convolutional neural network (CNN) is the most commonly used deep learning architecture in
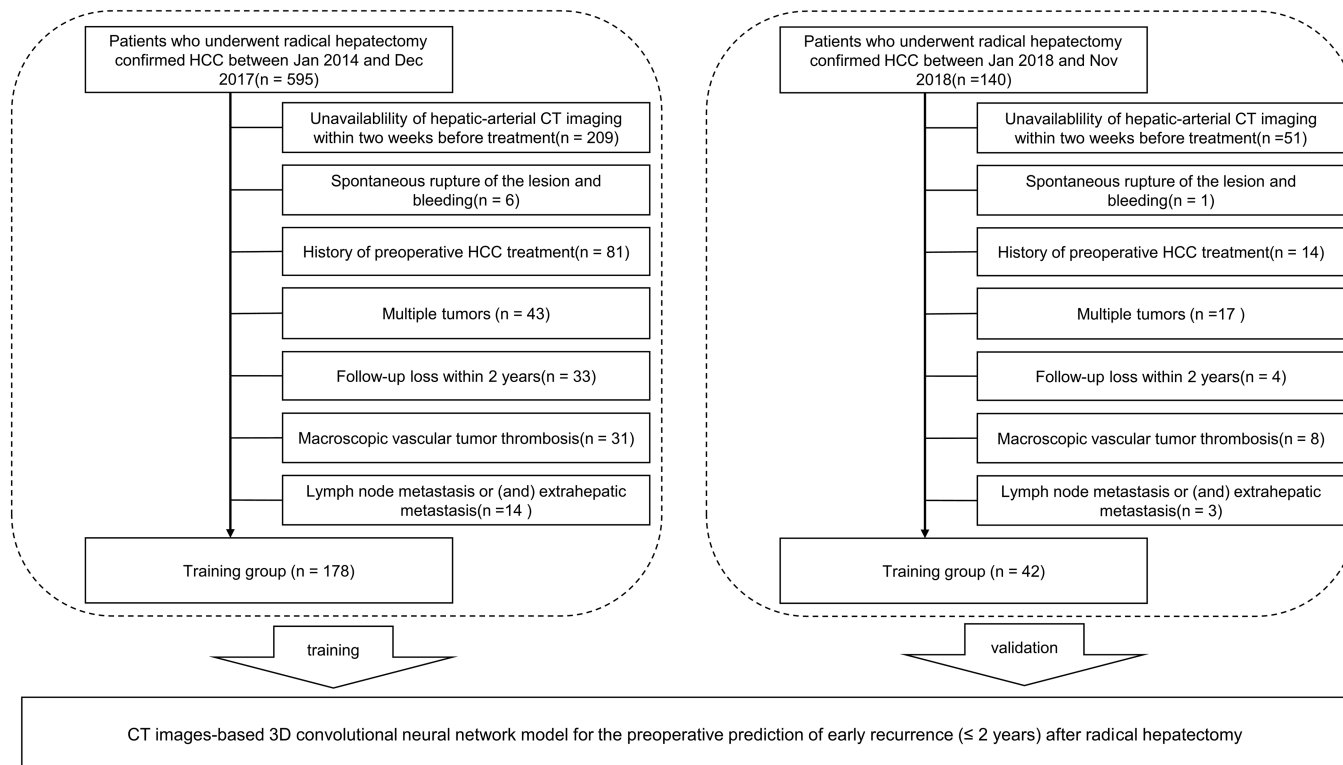
Figure 1. The flowchart.

medical image analysis.[13] The CNN model can predict the results directly from images without extracting artificially designed features. Therefore, CNNs have the ability to identify critical features that may be missing. Among them, 3-dimensional (3D) CNN can take advantage of both 1-dimensional and 2-dimensional (2D) CNNs by extracting both spectral and spatial features simultaneously from the input volume. This is particularly helpful for analyzing volumetric data in medical imaging, including CT, MR, and PET images.[14] Three-dimensional CNN has been applied in medical image classification, segmentation, localization, and detection. Zhou et al.[15] proposed a novel fully CNN constructed by 3D atrous-convolution for fully automatic brain tumor segmentation via MR images. Gu et al.[16] presented a 3D deep CNN-based framework with multiscale nodule prediction for automatic lung-nodule detection in CT scans.

In the previous studies of CNN of medical images, we found that there are two sampling methods. One is segmentation sampling (SS) that segments the tumor region in detail and deletes the background without tumors.[17,18] The other is non-segmentation sampling (NSS) that does not segment the tumors region and retains all of the background information.[19-21] These two segmentation methods are widely used in studies but few studies have compared them.

In this study, we developed 3D-CNNs based on contrast-enhanced CT images for the preoperative prediction of ER of solitary HCC and compared the effects of SS and NSS on the prediction performance of 3D-CNN.

Methods

Patients

This study was approved by the Institutional Review Board and Ethical Committee (NFEC-201208-K3), and the requirement for informed consent was waived because of the retrospective study design. We reviewed the contrast-enhanced CT images of 735 patients who underwent radical hepatectomy in the period 2014-2018 and were pathologically proven HCC. The inclusion criteria were as follows: (a) a solitary lesion pathologically proven as HCC; (b) availability of hepatic-arterial CT imaging within 2 weeks before treatment; (c) received initial treatment of radical hepatectomy; and (d) negative surgical margin. The exclusion criteria were as follows: (a) history of preoperative HCC treatment; (b) other malignant tumors; (c) Child–Pugh class C; (d) macroscopic vascular tumor thrombosis or (and) extrahepatic metastasis; (e) spontaneous rupture of the lesion and bleeding; and (f) the patients without recurrence or metastasis who were followed up for less than 2 years. The flowchart is presented in Figure 1.

Finally, 178 patients (January 2014 to December 2017) with HCC constituted the training group (73 patients with ER). Forty-two patients (January 2018 to November 2018) were allocated to the test group (21 patients with ER). The first follow-up appointments were fixed 1-2 months after the surgery and every 3–4 months thereafter.

Main points

- The segmentation sampling-3-dimensional-convolutional neural network (SS-3D-CNN) successfully stratified patients according to the differences in recurrence-free survival.

- We plotted the gradient-weighted class activation mappings and found that non-segmentation sampling-3-dimensional-convolutional neural network (NSS-3D-CNN) compared with SS-3D-CNN was obviously interfered by the nearby tissues.

- NSS is better than SS in improving the performance of 3D-CNN in our study.

CT acquisition and volume-of-interest segmentation

Supplementary S1 and Table S1 presents the scan characteristics and image acquisition process. In the original CT image, the resolution of the x-axis and y-axis ranged from 0.55859 to 0.94727 mm/pixel, and the resolution of the z-axis ranged from 5 to 8 mm/pixel. Therefore, a linear interpolation algorithm was applied for resampling to normalize the resolution to 1 × 1 × 1 mm/pixel and make images of different resolutions comparable in size and shape.

The hepatic-arterial CT images of all patients were fed into ITK-SNAP (version 3.8.0). The SS-volumes-of-interest (VOIs) in training group were analyzed and manually segmented by 2 abdominal CT interpretation radiologists. Disagreements between the radiologists about delineating the VOI were resolved by discussion and consensus. The SS-VOIs in test group were manually segmented by a liver oncologist. On the level, a region of 150 × 150 pixels that could cover the whole tumor was selected. For larger tumors, a height of 10 pixels was necessary in the center of the tumor. Therefore, the SS-VOIs were determined to be 3D patches of size 10 × 150 × 150 pixels (Figure 2b). Voxels without tumors were set to 0. The NSS-VOIs were determined by marking the bounding box of tumors without manually segmenting in detail (Figure 2c).

Data pre-processing

Data augmentation was used to increase the diversity of the training data set, prevent model overfitting, and improve the robustness of the model.[22] In the training group, the SS-VOIs and NSS-VOIs were expanded 7 times by data augmentation techniques, including up-down flip, left-right flip, random translation, 90° rotation, 180° rotation, and 270° rotation. This augmentation was not performed in the test group.

Model structure

We created a 3D-CNN model in python (version 3.8) using Keras framework (version 2.3.1). In designing the 3D-CNN, we use small convolution kernel and continuous convolution layers which may be more effective in the previous study,[23] and the 3D-CNN should not be too deep to avoid overfitting due to the limited size of the training group. Dropout was applied to reduce complex co-adaptations of neurons,
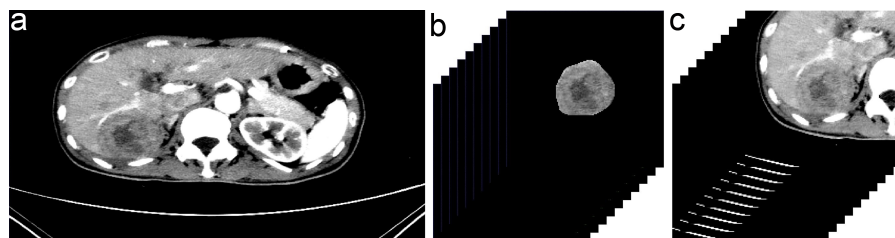
Figure 2. a-c. Segmentation technique. **(a)** A hepatic-arterial CT; **(b)** a SS-VOI; **(c)** a NSS-VOI. CT, computed tomography, SS, segmentation sampling; NSS, non-segmentation sampling; VOI, volume-of-interest.
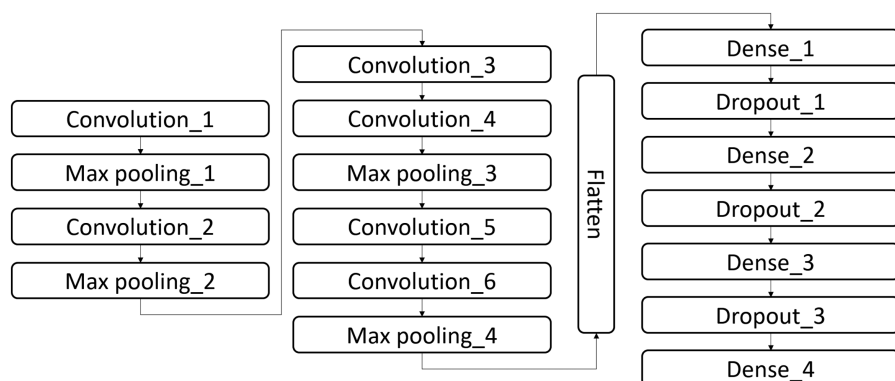
Figure 3. Model structure.
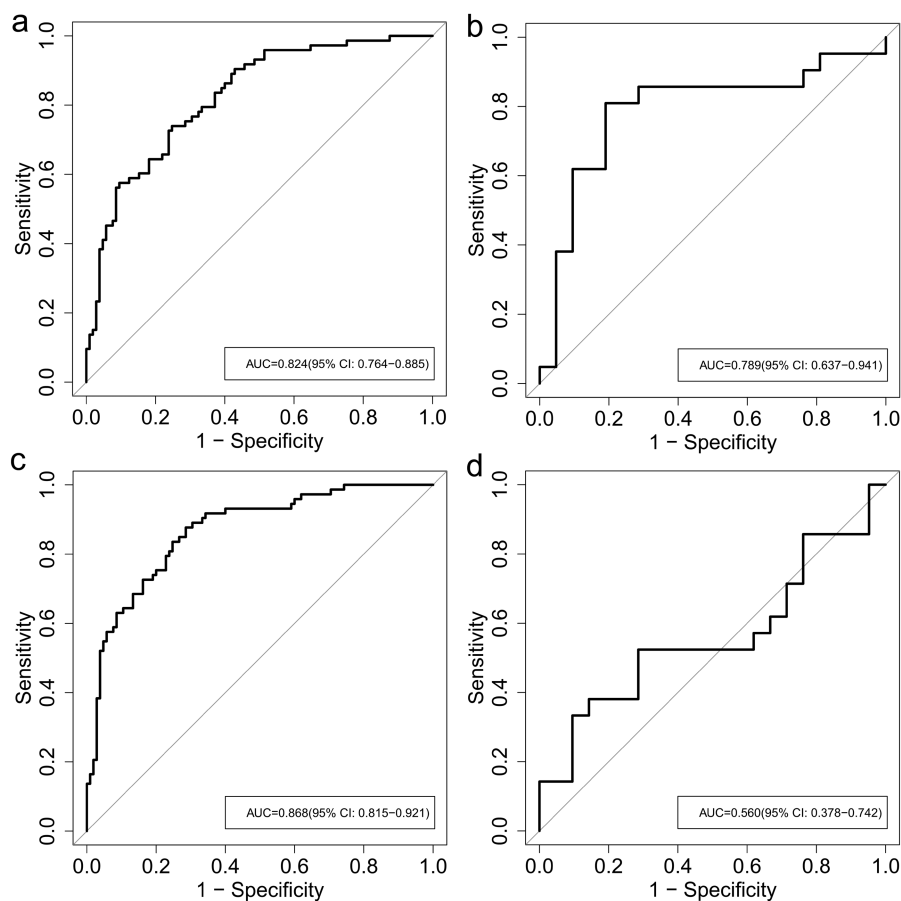
Figure 4. a-d. The ROCs of the 3D-CNNs. **(a)** The ROC of SS-3D-CNN in the training group; **(b)** the ROC of SS-3D-CNN in the test group; **(c)** the ROC of NSS-3D-CNN in the training group; **(d)** the ROC of NSS-3D-CNN in the test group. ROC, receiver operating characteristic; CNN, convolutional neural network.

which consisted of setting the output of each hidden neuron with a probability of 0.5 to 0.[24] The structure of the model is shown in Figure 3. The detailed parameters of the model can be seen in Supplementary Table S2.

Statistical analysis

For statistical analysis, we plotted the receiver operating characteristic (ROC) curves, the calibration curves, and calculated the area under the curves (AUCs) to evaluate the SS-3D-CNN and NSS-3D-CNN for discriminating ER and non-ER patients. Survival curves were generated using the Kaplan–Meier method and compared using a 2-sided log-rank test ($P < .001$ was considered statistically significant). The gradient-weighted class activation mappings (Grad-CAMs) were plotted using keras-vis (version 0.4.1). Risk factors of ER were identified using univariate analysis. The chi-square test or Fisher exact test was used to compare categorical variables where appropriate ($P < .05$ was considered statistically significant).

Training

In the training process, the 5-fold cross-validation was applied. In each round of cross-validation, 80% of the VOIs of the training group were used as training data; the remaining 20% were used as validation data. Because the ER and non-ER (NER) patients were not evenly distributed in the training group, we balanced the weights of
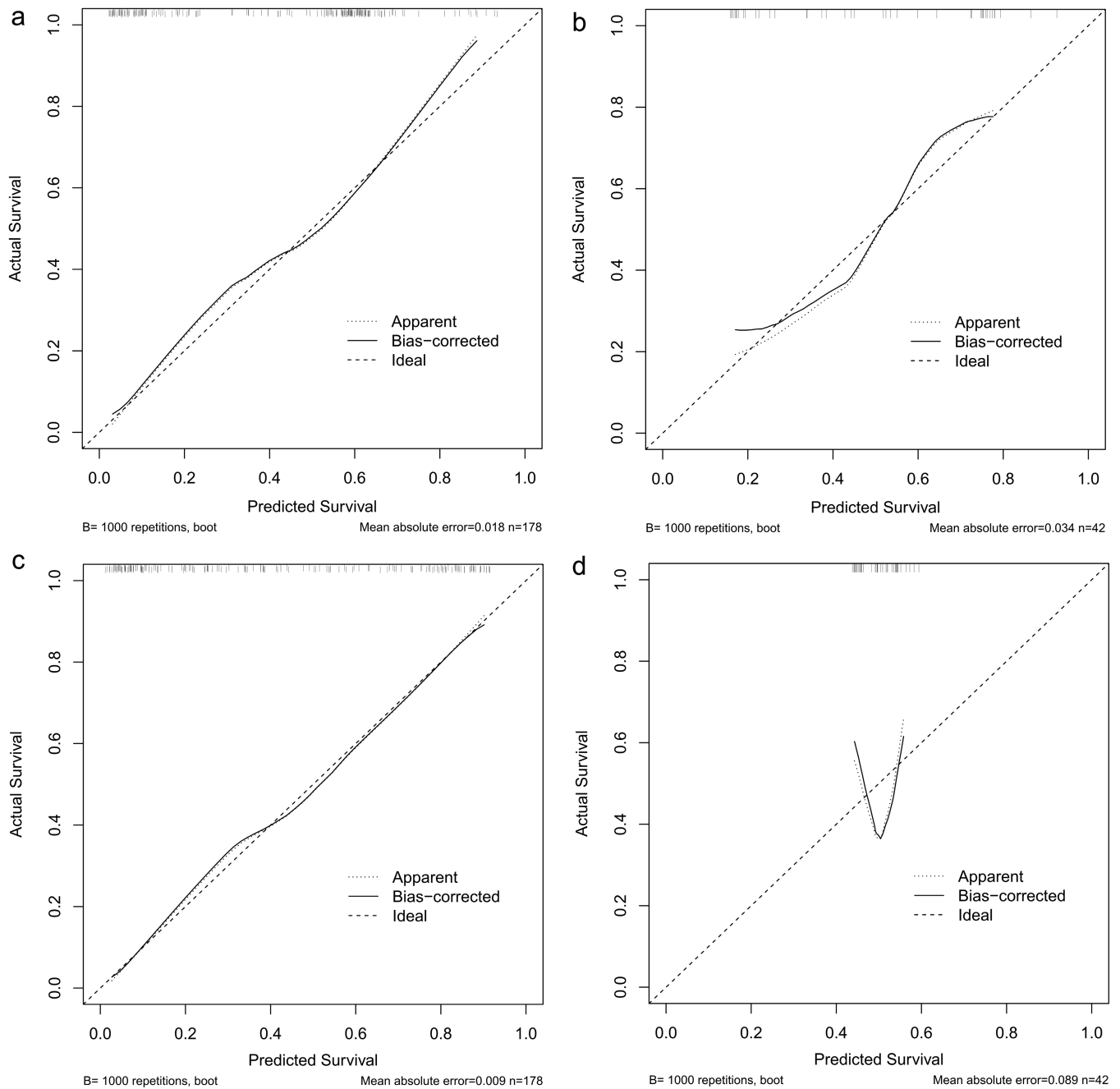
Figure 5. a-d. The calibration curves of the 3D-CNNs. **(a)** The calibration curve of SS-3D-CNN in the training group; **(b)** the calibration curve of SS-3D-CNN in the test group; **(c)** the calibration curve of NSS-3D-CNN in the training group; **(d)** the calibration curve of NSS-3D-CNN in the test group.

the 2 sets of data to improve the accuracy of the model. The model was optimized by Adam with a global learning rate of 0.00001 without decay. The SS-VOIs and NSS-VOIs were respectively used to train SS-3D-CNN and NSS-3D-CNN. The training experiments were performed on a machine with the following specifications: Intel(R) Core (TM) i5-10300H CPU @ 2.50 GHz 64-bit, GPU NVIDIA GeForce GTX 1650, 4GB of RAM, and the Windows 10 Home version 1909 system.

Results

The baseline characteristics of patients in the training and test groups are presented in Supplementary Table S3. The statistical information of the data is presented in Supplementary Table S4. No information rates in training group and test group are 0.60 and 0.50. We tested the SS-3D-CNN model using the training and test groups (training group: AUC of 0.824 (95% CI: 0.764-0.885), Figure 4a; test group: AUC of 0.789 (95% CI: 0.637-0.941), Figure 4b) and tested NSS-3D-CNN in the same way (training group: AUC of 0.868 (95% CI: 0.0.815-0.921), Figure 4c; test group: AUC of 0.560 (95% CI: 0.378-0.742), Figure 4d). The calibration curves of 2 models are shown in Figure 5. In Supplementary Figure S1, we compared the performance of the CNN for 2 different raters who do the segmentations and found that the results are similar.

The confusion matrixes of SS-3D-CNN and NSS-3D-CNN are shown in Figure 6. For the test group, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy value of SS-3D-CNN were 61.9%, 90.5%, 86.7%, 70.4%, and 0.762, respectively (95% CI: 0.606, 0.880) ($P < .001$). For the test group, sensitivity, specificity, PPV, NPV, and accuracy value of NSS-3D-CNN were 52.4%, 61.9%, 57.9%, 56.5%, and 0.571, respectively (95% CI: 0.410, 0.723) ($P = .2204$).

Based on the maximum Youden index, we used AUC to determine the optimal cutoff value (SS-3D-CNN: 0.598; NSS-3D-CNN: 0.503) based on the training group. The optimal cutoff value was applied to the training and test groups to separate
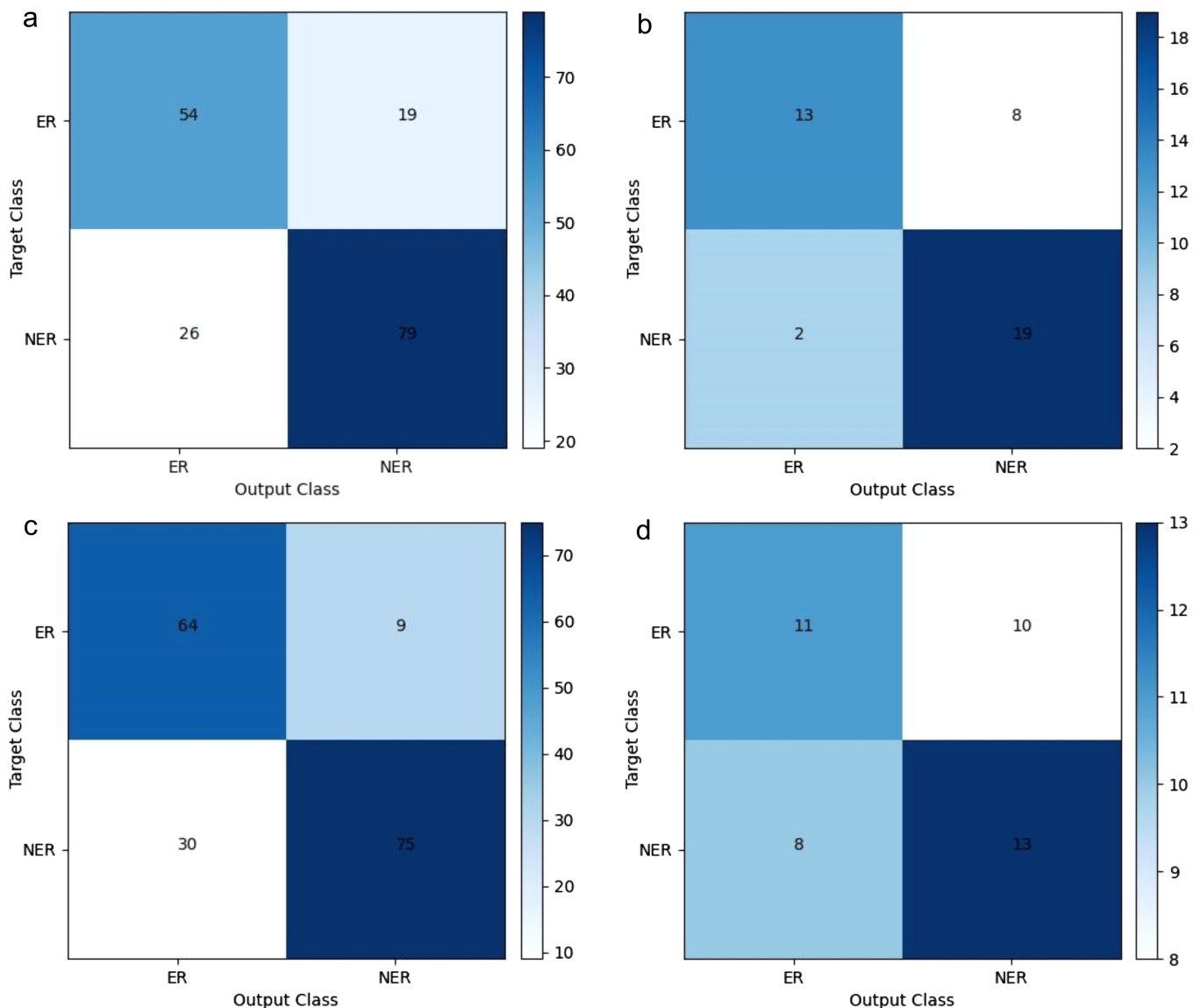
Figure 6. a-d. The confusion matrixes of the 3D-CNNs. **(a)** The confusion matrix of SS-3D-CNN in the training group; **(b)** the confusion matrix of SS-3D-CNN in the test group; **(c)** the confusion matrix of NSS-3D-CNN in the training group; **(d)** the confusion matrix of NSS-3D-CNN in the test group.
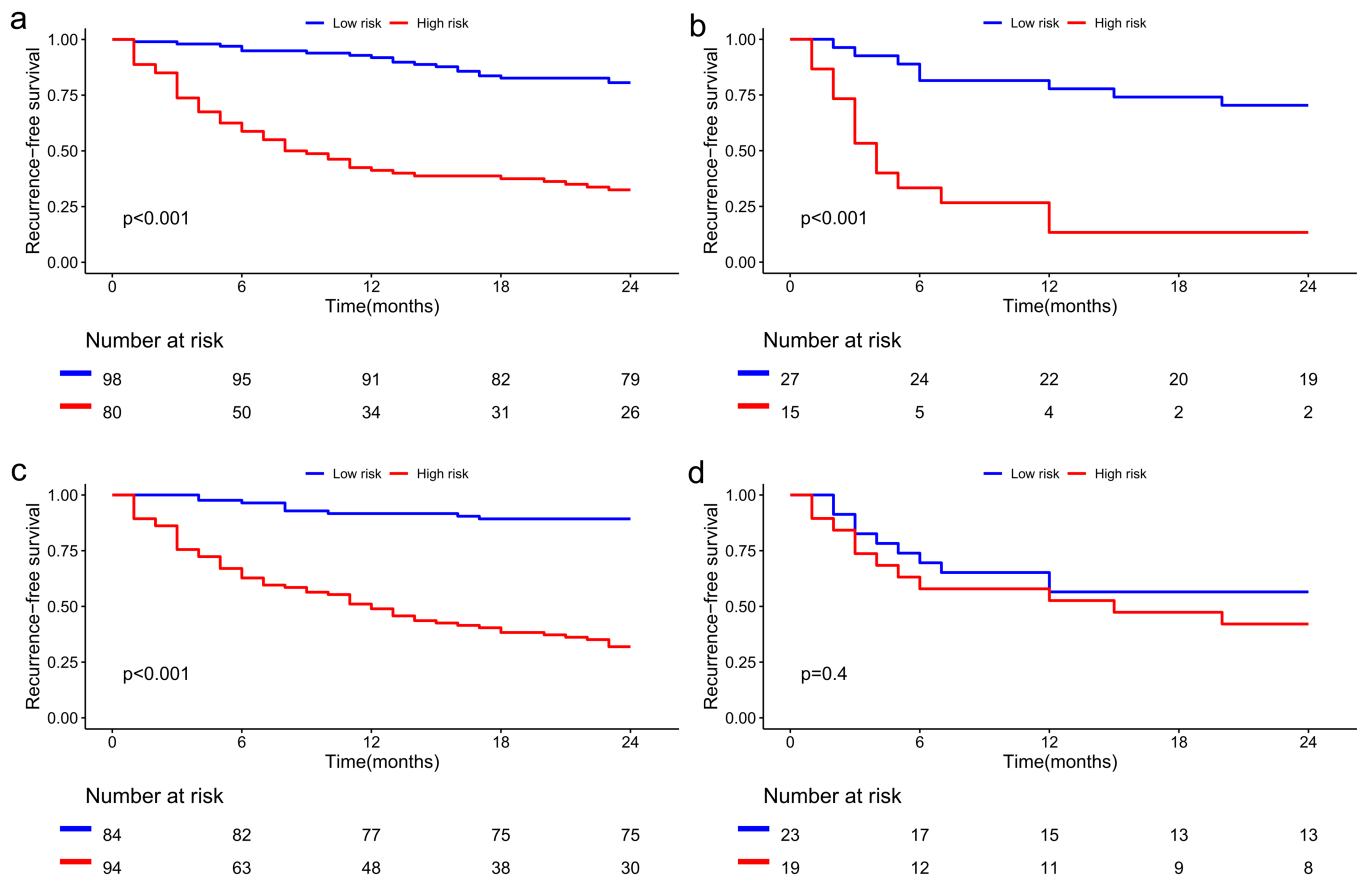
Figure 7. a-d. The Kaplan–Meier curves of RFS. **(a)** The Kaplan–Meier curve of SS-3D-CNN in the training group; **(b)** the Kaplan–Meier curve of SS-3D-CNN in the test group; **(c)** the Kaplan–Meier curve of NSS-3D-CNN in the training group; **(d)** the Kaplan–Meier curve of NSS-3D-CNN in the test group. RFS, recurrence-free survival.

patients into low- and high-risk groups. The Kaplan–Meier curves of recurrence-free survival (RFS) of the SS-3D-CNN and NSS-3D-CNN in the 2 groups are shown in Figure 7. The Kaplan–Meier curve of SS-3D-CNN in the test group demonstrates a significant difference in RFS ($P \leq .001$) (in Figure 7b) but that of NSS-3D-CNN demonstrates a non-significant difference in RFS ($P = .4$) (in Figure 7d). In the training group, the median RFS of high-risk groups in the SS-3D-CNN and NSS-3D-CNN was 8 ± 1.79 and 12 ± 1.76 months and the median RFS of low-risk groups was not reached in both 3D-CNNs. In the test group, the median RFS of high-risk groups in the SS-3D-CNN and NSS-3D-CNN was 4 ± 0.76 and 15 ± 10.16 months and the median RFS of low-risk groups was not reached in both 3D-CNNs.

Next, we plotted Grad-CAMs to analyze the difference of prediction logic between the SS-3D-CNN and NSS-3D-CNN (in Figure 8a). Grad-CAMs showed that NSS-3D-CNN had the ability to localize tumors (in Figure 8b), but was obviously interfered by nearby tissues, while SS-3D-CNN focuses on the tumor area.

Discussion

In this study, we assessed the ability of 3D-CNN for the preoperative prediction of 2-year RFS of solitary HCC patients after radical hepatectomy based on contrast-enhanced CT data. According to different sampling methods, we have completed 2 different models, SS-3D-CNN and NSS-3D-CNN, which have the same model structure, in order to compare the impact of the 2 sampling methods on the prediction results. This SS-3D-CNN successfully stratified patients according to the differences in RFS and may be of clinical use in identifying optimal treatment strategies. Multimodality therapy is increasingly being explored to decrease the incidence of tumor recurrence; however, the effectiveness of multimodality therapy is controversial.[25] We speculate that this may be related to the lack of effective stratification of patients; 3D-CNN may be an effective tool for stratification.

According to ROC, calibration curves, confusion matrixes, and the Kaplan–Meier curves in the test group, the generalization performance of SS-3D-CNN is obviously better than that of NSS-3D-CNN. In order to explore the reason for the poor performance of NSS-3D-CNN and the difference of prediction logic between 2 sampling methods, we plotted Grad-CAMs which shows that NSS-3D-CNN is interfered by nearby tissues. SS forces the model to focus on the tumor area by deleting the background without tumor, but SS needs more human resources and time to segment the tumor manually. NSS retains the background information and economizes in time and human resources, but excessive background information also interferes with the model, especially when the sample size is limited.

There are previous studies that have used CNN to predict the prognosis of tumors. Hosny et al.[20] trained 3D-CNN model to predict 2-year survival in radiotherapy patients (AUC = 0.70) and surgical patients (AUC = 0.71) with non-small cell lung cancer. In their study, the researchers used NSS to get VOIs, but they got good results. Therefore, we speculate that non-small cell lung cancer is different from HCC and its boundary with surrounding tissues is
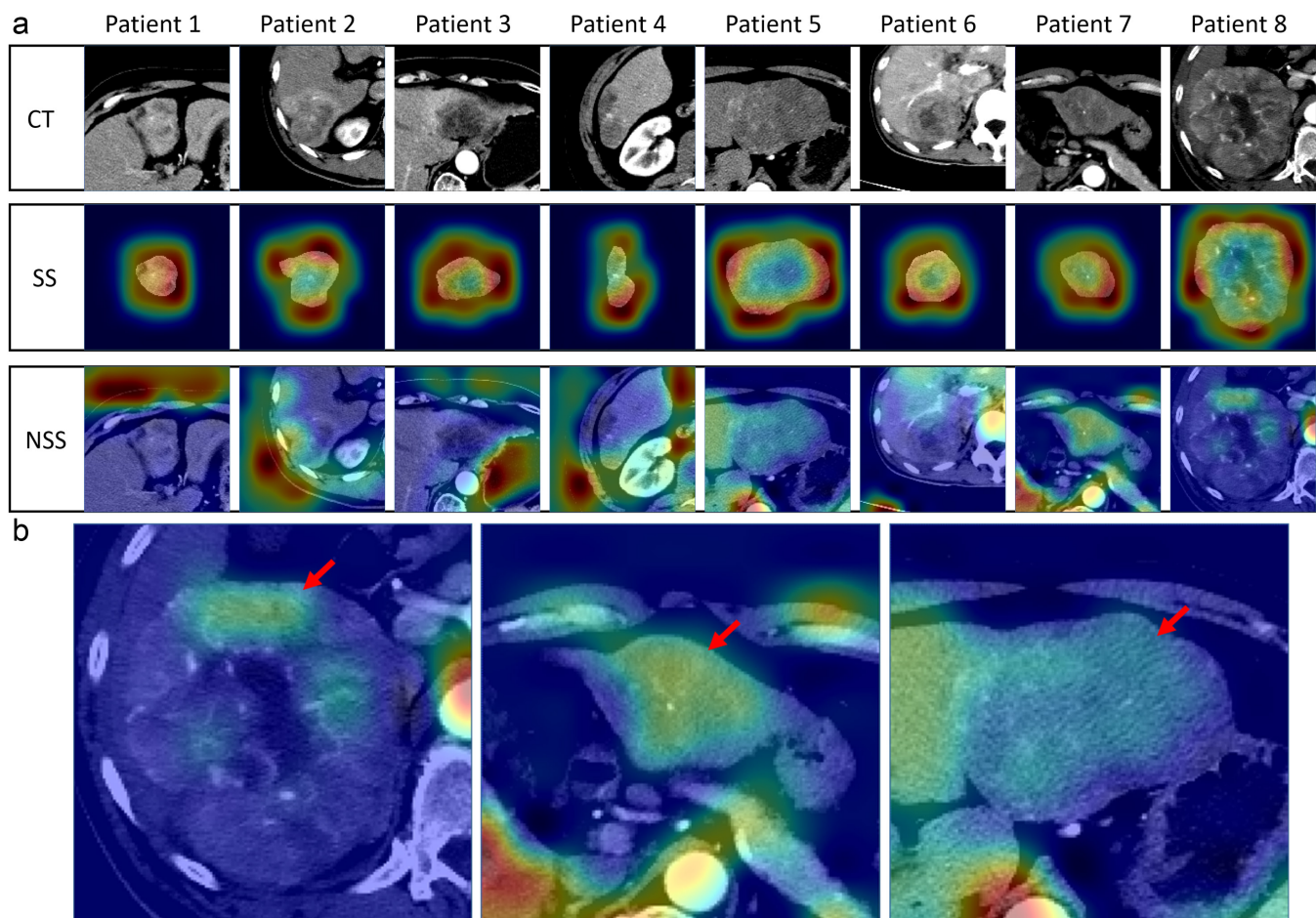
Figure 8. a-b. The Grad-CAMs of the 3D-CNNs. **(a)** Comparison of the differences of the Grad-CAMs of SS-3D-CNN and NSS-3D-CNN; **(b)** NSS-3D-CNN have the ability to localize tumors, but it was also obviously interfered by nearby tissues (the *red arrows* refer to the tumors area that the model focuses on). Note: In our study, the penultimate layer is the max pooling 4 after all the convolution layers. So the Grad-CAMs are 2D images. Grad-CAM, gradient-weighted class activation mapping.

clearer, which is conducive to the correct recognition of the model. And they made all-inclusive sampling in surgical patients. In our study, the size of the tumors is large and the size span between the tumors is large, too. So it is difficult for us to determine a fixed volume to achieve all-inclusive sampling. Jiang et al.[19] developed a 3D-CNN model to assess microvascular invasion in HCC patients, and the AUC of the 3D-CNN model in the training set and the validation set were 0.980 and 0.906, respectively. Next, the researchers found that this model also stratified patients in RFS. In this article, they also used NSS to get VOIs, which can get good results in assessing microvascular invasion. But SS-3D-CNN in our study is better than their model in terms of stratifying patients. A recent study built deep learning model based on 2D imagings to predict the ER of HCC. They developed the clinical model (AUC = 0.7532) and the deep learning model (AUC = 0.7233) and proposed 4 fusion models to combine clinical data and deep learning features which achieved higher AUC.[21] However, they did not set up test group, which made it difficult to evaluate the generalization ability of the fusion models.

In this study, 2 major drawbacks affect the performance of our 3D-CNN model. First, the sample size was relatively small, which easily caused the curse of dimensionality, leading to overfitting. We used data augmentation techniques in the data pre-processing stage and added the dropout to the model structure to avoid overfitting. However, the relatively small sample size was still the major drawback that restricted the development of the model. In addition, the limited sample size greatly affects the performance of NSS-3D-CNN. In the future, we will try our best to seek a much larger database of the prospective study would be collected from more centers to further compare the impact of SS and NSS on the prediction performance of 3D-CNN. Second, the high computational cost of the 3D-CNN limited the size of the VOI. Compared with 2D-CNN, the high computational cost of 3D-CNN is the main obstacle to its development. Thus, when the tumor diameter is very large, the VOI cannot contain the complete tumor. The problem needs to be further studied in future.

In conclusion, it is feasible to use CT images-based 3D-CNN models to predict ER of HCC. Additionally, SS can help 3D-CNN get better performance than NSS when the sample size is limited.

Author Contributions

(I) Conception and design: Hao Cui, Kun-Yuan Wang, Wen-Yuan Li; (II) Administrative support: Li Liu; (III) Provision of study material or patients: Li Liu; (IV) Collection and assembly of data: Hao Cui, Kun-Yuan Wang, Wen-Yuan Li, Hong-Bo Zhu, Lu-Shan Xiao; (V) Data analysis and interpretation: Hao Cui, Kun-Yuan Wang, Wen-Yuan Li, Hong-Bo Zhu; (VI) Manuscript writing: All

authors; (VII) Final approval of manuscript: All authors.

References

1. Yang JD, Hainaut P, Gores GJ, Amadou A, Plymoth A, Roberts LR. A global view of hepatocellular carcinoma: trends, risk, prevention and management. *Nat Rev Gastroenterol Hepatol*. 2019;16(10):589-604. [CrossRef]
2. Xie DY, Ren ZG, Zhou J, Fan J, Gao Q. Chinese clinical guidelines for the management of hepatocellular carcinoma: updates and insights. *Hepatobiliary Surg Nutr*. 2020;9(4):452-463. [CrossRef]
3. European Association for the Study of the Liver. Electronic address: easloffice@easloffice.eu, European Association for the Study of the Liver. EASL Clinical Practice Guidelines. EASL Clinical Practice Guidelines: management of hepatocellular carcinoma. *J Hepatol*. 2018; 69(1):182-236. [CrossRef]
4. Yang JD, Heimbach JK. New advances in the diagnosis and management of hepatocellular carcinoma. *BMJ*. 2020;371:m3544. [CrossRef]
5. Llovet JM, Fuster J, Bruix J. Intention-to-treat analysis of surgical treatment for early hepatocellular carcinoma: resection versus transplantation. *Hepatology*. 1999;30(6):1434-1440. [CrossRef]
6. Kulik L, El-Serag HB. Epidemiology and management of hepatocellular carcinoma. *Gastroenterology*. 2019;156(2):477-491.e1. [CrossRef]
7. Llovet JM, Villanueva A, Marrero JA, et al. Trial design and endpoints in hepatocellular carcinoma: AASLD consensus conference. *Hepatology*. 2021;73(suppl 1):158-191. [CrossRef]
8. Zhu HB, Zheng ZY, Zhao H, et al. Radiomics-based nomogram using CT imaging for noninvasive preoperative prediction of early recurrence in patients with hepatocellular carcinoma. *Diagn Interv Radiol*. 2020;26(5):411-419. [CrossRef]
9. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. 2012;48(4):441-446. [CrossRef]
10. Gillies RJ, Kinahan PE, Hricak HH. Radiomics: Images are more than pictures, they are data. *Radiology*. 2016;278(2):563-577. [CrossRef]
11. Yan J, Zhang S, Li KK, et al. Incremental prognostic value and underlying biological pathways of radiomics patterns in medulloblastoma. *EBiomedicine*. 2020;61:103093. [CrossRef]
12. Park JE, Ham S, Kim HS, et al. Diffusion and perfusion MRI radiomics obtained from deep learning segmentation provides reproducible and comparable diagnostic model to human in post-treatment glioblastoma. *Eur Radiol*. 2021;31(5):3127-3137. [CrossRef]
13. Zhou LQ, Wang JY, Yu SY, et al. Artificial intelligence in medical imaging of the liver. *World J Gastroenterol*. 2019;25(6):672-682. [CrossRef]
14. Singh SP, Wang L, Gupta S, Goli H, Padmanabhan P, Gulyás B. 3D deep learning on medical images: a review. *Sensors (Basel)*. 2020;20(18):5097. [CrossRef]
15. Zhou Z, He Z, Jia Y. AFPNet: A 3D fully convolutional neural network with atrous-convolution feature pyramid for brain tumor segmentation via MRI images. *Neurocomputing*. 2020; 402:235-244. [CrossRef]
16. Gu Y, Lu X, Yang L, et al. Automatic lung nodule detection using a 3D deep convolutional neural network combined with a multi-scale prediction strategy in chest CTs. *Comput Biol Med*. 2018;103:220-231. [CrossRef]
17. Peng J, Kang S, Ning Z, et al. Residual convolutional neural network for predicting response of transarterial chemoembolization in hepatocellular carcinoma from CT imaging. *Eur Radiol*. 2020;30(1):413-424. [CrossRef]
18. Zhou Q, Zhou Z, Chen C, et al. Grading of hepatocellular carcinoma using 3D SE-DenseNet in dynamic enhanced MR images. *Comput Biol Med*. 2019;107:47-57. [CrossRef]
19. Jiang YQ, Cao SE, Cao S, et al. Preoperative identification of microvascular invasion in hepatocellular carcinoma by XGBoost and deep learning. *J Cancer Res Clin Oncol*. 2021;147(3):821-833. [CrossRef]
20. Hosny A, Parmar C, Coroller TP, et al. Deep learning for lung cancer prognostication: A retrospective multi-cohort radiomics study. *PLOS Med*. 2018;15(11):e1002711. [CrossRef]
21. Wang W, Chen Q, Iwamoto Y, et al. Deep fusion models of multi-phase CT and selected clinical data for preoperative prediction of early recurrence in hepatocellular carcinoma. *IEEE Access*. 2020;8:139212-139220. [CrossRef]
22. Huang L, Pan W, Zhang Y, Qian L, Gao N, Wu Y. Data augmentation for deep learning-based radio modulation classification. *IEEE Access*. 2020;8:1498-1506. [CrossRef]
23. Tran D, Bourdev L, Fergus R, Torresani L, Paluri M. Learning spatiotemporal features with 3D convolutional networks. In: *Proceedings of the IEEE International Conference on Computer Vision 2015*. ICCV; 2015:4489-4497. [CrossRef]
24. Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. *Insights Imaging*. 2018;9(4):611-629. [CrossRef]
25. Akateh C, Black SM, Conteh L, et al. Neoadjuvant and adjuvant treatment strategies for hepatocellular carcinoma. *World J Gastroenterol*. 2019;25(28):3704-3721. [CrossRef]
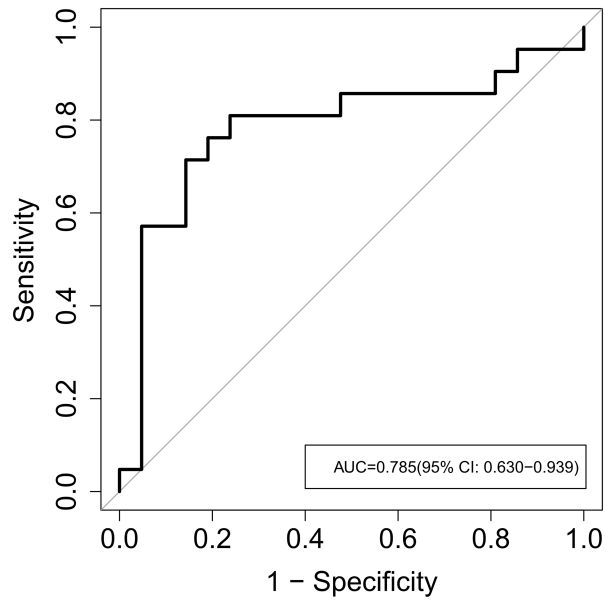
Supplementary Figure S1. The ROC of SS-VOIs which were segmented by a radiology expert. We invited a radiologist to independently segment the SS-VOIs in test group. The AUC of the SS-VOIs segmented by the radiologist in the test group was 0.785 (95% CI: 0.630-0.939) (in the figure), which is similar with AUC of the SS-VOIs segmented by the liver oncologist.

All patients from the Nanfang Hospital, Southern Medical University underwent contrast-enhanced CT (CECT) using either of two multi-detector row CT (MDCT) systems: the SOMATOM (Siemens Medical Systems) or the Brilliance iCT 256 (Philips Healthcare). The scan characteristics are listed in Supplementary Table 1.

Additionally, we injected contrast material (1.5 mL/kg, Ultravist 370, Bayer Schering Pharma) intravenously at a flow rate of 2.0–3.0 mL/s using a pump injector (Ulrich CT Plus 150, Ulrich Medical) to obtain CECT images. Four-phase (unenhanced, hepatic-arterial, portal venous, and delayed phases) CT images were obtained at 0 s, 30 s, 60 s, and 120 s after injection, respectively.

Supplementary Table S1. The scan characteristics		
Parameter	SOMATOM	Brilliance iCT256
Tube voltage (kVp)	120	120
Tube current (mA)	Auto	Auto
Detector collimation (mm)	64×0.6	128×0.625
Field of view (mm)	250-500	300-400
Matrix size	512×512	512×512
Rotation times (s)	0.5	0.5
Slice interval (mm)	0	0
Slice thickness (mm)	1-5	1-5

Supplementary Table S2. The detailed parameters of 3D-CNN

Layer	Kernel	Kernel size	Strides	Activation	Padding	Output shape
Convolution1	32	3, 3, 3	1, 1, 1	Relu	Same	10, 150, 150, 32
Max pooling1	-	3, 3, 3	2, 2, 2	-	Same	5, 75, 75, 32
Convolution2	64	3, 3, 3	1, 1, 1	Relu	Same	5, 75, 75, 64
Max pooling2	-	3, 3, 3	2, 2, 2	-	Same	3, 38, 38, 64
Convolution3	128	3, 3, 3	1, 1, 1	Relu	Same	3, 38, 38, 128
Convolution4	128	3, 3, 3	1, 1, 1	Relu	Same	3, 38, 38, 128
Max pooling3	-	3, 3, 3	2, 2, 2	-	Same	2, 19, 19, 128
Convolution5	256	2, 3, 3	1, 1, 1	Relu	Same	2, 19, 19, 256
Convolution6	256	2, 3, 3	1, 1, 1	Relu	Same	2, 19, 19, 256
Max pooling4	-	2, 3, 3	2, 2, 2	-	Same	1, 10, 10, 256
Flatten	-	-	-	-	-	25600
Dense1	512	-	-	Relu	-	512
Dropout1	-	-	-	-	-	512
Dense2	256	-	-	Relu	-	256
Dropout2	-	-	-	-	-	256
Dense3	128	-	-	Relu	-	128
Dropout3	-	-	-	-	-	128
Dense4	2	-	-	Softmax	-	2

Supplementary Table S3. The baseline characteristics of patients in the training and test groups			
Characteristic	Training group (n = 178)	Test group (n = 42)	P
Age (years)			
<60	139	29	.215
≥60	39	13	
Sex			
Male	157	36	.658
Female	21	6	
HBsAg status			
Positive	160	35	.277
Negative	18	7	
Child–Pugh classification			
A	162	37	.563
B	16	5	
Hepatocirrhosis status			
Present	111	25	.734
Absent	67	17	
Tumor size (cm)			
≤5	86	19	.720
>5	92	23	
Pathological grade			
Poorly differentiated	28	6	.124
Moderately differentiated	127	35	
Well differentiated	23	1	
ALT (U/mL)			
≤40	132	27	.199
>40	46	15	
AST (U/mL)			
≤40	157	33	.102
>40	21	9	
AFP (ng/mL)			
≤20	80	18	.807
>20	98	24	
Class			
ER	73	21	.289
NER	105	21	

ALT, alanine aminotransferase; AST, aspartate aminotransferase; AFP, serum alpha-fetoprotein; ER, Early recurrence; NER, Non-early-recurrence.

Supplementary Table S4. The statistical information of the data			
Characteristic	NER (n=105)	ER (n = 73)	P
Age (years)			
<60	81	58	.714
≥60	24	15	
Sex			
Male	93	64	.855
Female	12	9	
HBsAg status			
Positive	93	67	.485
Negative	12	6	
Child–Pugh classification			
A	100	62	.018
B	5	11	
Hepatocirrhosis status			
Present	62	49	.274
Absent	43	24	
Tumor size (cm)			
≤5	64	22	<.001
>5	41	51	
Pathological grade			
Poorly differentiated	16	3	.043
Moderately differentiated	69	58	
Well differentiated	20	12	
ALT (U/L)			
≤40	80	52	.457
>40	25	21	
AST (U/L)			
≤40	95	62	.259
>40	10	11	
AFP (ng/L)			
≤20	56	24	.007
>20	49	49	

ALT, alanine aminotransferase; AST, aspartate aminotransferase; AFP, serum alpha-fetoprotein.