Heliyon 10 (2024) e31451

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Deep learning model based on contrast-enhanced MRI for predicting post-surgical survival in patients with hepatocellular carcinoma

Lidi Ma [a,1], Congrui Li [b,1], Haixia Li [c,1], Cheng Zhang [a], Kan Deng [d], Weijing Zhang [a], Chuanmiao Xie [a,*]

[a] Department of Radiology, State Key Laboratory of Oncology in South China, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-Sen University Cancer Center, Guangzhou, 510060, PR China
[b] Department of Diagnostic Radiology, Hunan Cancer Hospital, Central South University, Changsha, PR China
[c] Bayer, Guangzhou, Guangdong, PR China
[d] Clinical Science, Philips Healthcare, Guangzhou, PR China

ARTICLE INFO

Keywords:
Deep learning
Carcinoma
Hepatocellular
Multiparametric magnetic resonance imaging
Prognosis

ABSTRACT

Objective: To develop a deep learning model based on contrast-enhanced magnetic resonance imaging (MRI) data to predict post-surgical overall survival (OS) in patients with hepatocellular carcinoma (HCC).

Methods: This bi-center retrospective study included 564 surgically resected patients with HCC and divided them into training (326), testing (143), and external validation (95) cohorts. This study used a three-dimensional convolutional neural network (3D-CNN) ResNet to learn features from the pretreatment MR images (T1WIpre, late arterial phase, and portal venous phase) and got the deep learning score (DL score). Three cox regression models were established separately using the DL score (3D-CNN model), clinical features (clinical model), and a combination of above (combined model). The concordance index (C-index) was used to evaluate model performance.

Results: We trained a 3D-CNN model to get DL score from samples. The C-index of the 3D-CNN model in predicting 5-year OS for the training, testing, and external validation cohorts were 0.746, 0.714, and 0.698, respectively, and were higher than those of the clinical model, which were 0.675, 0.674, and 0.631, respectively ($P = 0.009$, $P = 0.204$, and $P = 0.092$, respectively). The C-index of the combined model for testing and external validation cohorts was 0.750 and 0.723, respectively, significantly higher than the clinical model ($P = 0.017$, $P = 0.016$) and the 3D-CNN model ($P = 0.029$, $P = 0.036$).

Conclusions: The combined model integrating the DL score and clinical factors showed a higher predictive value than the clinical and 3D-CNN models and may be more useful in guiding clinical treatment decisions to improve the prognosis of patients with HCC.

Abbreviations: AFP, alpha-fetoprotein; AIC, Akaike information criterion; ALB, albumin; AP, late arterial phase; BCLC, Barcelona Clinic Liver Cancer; C-index, Concordance index; CNN, convolutional neural network; DL, deep learning; DP, delayed phase; DWI, diffusion-weighted imaging; GD-DTPA, gadolinium-diethylene triamine pentaacetate; HCC, hepatocellular carcinoma; HR, hazard ratios; IQR, interquartile range; MVI, microvascular invasion; OS, overall survival; PVP, portal venous phase; ResNet, residual neural network; ResNet-18, 18-layer ResNet model; SD, standard deviation; T1WI, T1-weighted imaging; T1WIpre, precontrast T1WI; VOIs, volumes of interest.

* Corresponding author.
E-mail address: xchuanm@sysucc.org.cn (C. Xie).

[1] Lidi Ma, Congrui Li, and Haixia Li contributed equally to this work.

https://doi.org/10.1016/j.heliyon.2024.e31451
Received 22 October 2023; Received in revised form 15 May 2024; Accepted 16 May 2024
Available online 16 May 2024

Critical relevance statement

The combined model integrating the deep learning score and clinical factors showed a better predictive value than the clinical or three-dimensional convolutional neural network models.

1Introduction

Liver cancer is the sixth most common malignancy and the third leading cause of cancer-related deaths globally and accounts for approximately 90 % of hepatocellular carcinoma (HCC) cases [1], mainly due to the prevalence of viral hepatitis, aflatoxin B1, and alcohol intake [2]. There are various treatments for HCC, however, surgical resection is the recommended curative treatment for patients with resectable HCC at any stage [3]. Nevertheless, the 5-year survival rate of patients is still low at approximately 18–20 % [4], mainly due to the lack of regular and effective prognostic monitoring and follow-up mechanisms [5]. Therefore, it is critical to establish an effective prognostic monitoring model for patients with HCC after surgery to improve their prognosis.

Some clinical factors, such as the Barcelona Clinic Liver Cancer (BCLC) stage, microvascular invasion (MVI), serum alpha-fetoprotein (AFP), number or size of the tumor(s), hypointensity in the hepatobiliary phase, and incomplete capsule, are associated with postoperative prognosis [6–9]. Additionally, various nomograms based on staging systems have been developed to predict HCC outcomes [10]. Whereas, most of these studies have subjective interpretations and inconsistent criteria with low predictive performance (ROC = 0.62–0.69 for different staging systems) [9,11]. Moreover, most patients in these studies had HBV infection with a single institute, which means that patients with different ethnic groups and hepatitis backgrounds need to be further validated. Besides, some of the risk factors can only be obtained through postoperative pathology. There have been several attempts to quantitatively characterize the prognosis of HCC in recent years. The survival of patients with HCC may be predicted preoperatively by computed tomography (CT) texture analysis [12]. Compared with CT, dynamic contrast-enhanced magnetic resonance imaging (MRI) provides high tissue resolution, multiple parameters, and function information [13,14]. An MRI-based radiomics model was developed by Wang et al. [15] in order to predict patients' prognosis in HCC. Non-invasive accurate predictions of patient survival through preoperative MRI information will be of great significance in enhancing the surveillance of high-risk patients and guiding clinical decision-making. However, radiomics has its limitations; it can only extract specific and superficial features, many of which tend to be redundant.

Deep learning (DL), a subfield of machine learning, has been widely used in medical image analysis [16]. Compared with the superficial features of radiomics, limited to image heterogeneity expression, a convolutional neural network (CNN) has several hidden layers associated with predefined tasks by nonlinear functions that learn features and is superior to radiomics in performance [17–19]. DL models can detect the biological characteristics of HCC and can be used for liver focal lesion segmentation [20], differentiation between benign and malignant liver tumors, fibrosis staging [21,22], MVI prediction [23], and prognostic predictions of patients after surgical resection [24], among others, with good predictive performance. However, studies predicting the overall survival (OS) of

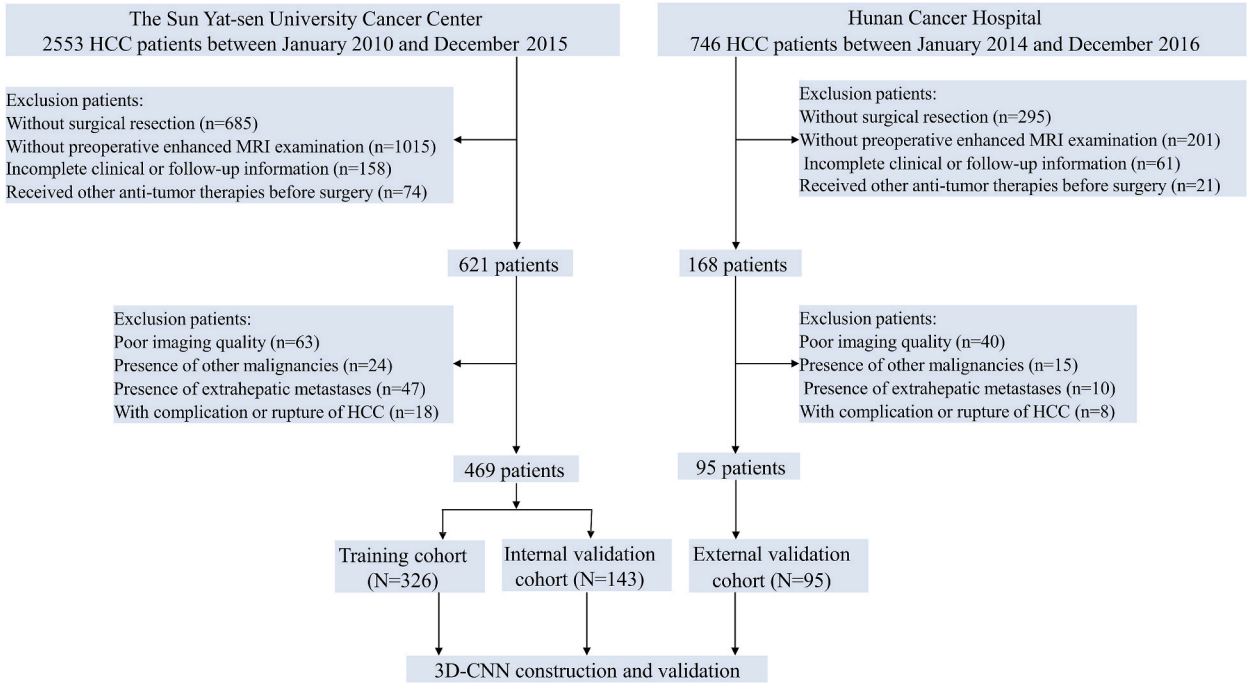

Fig. 1. Flowchart of the study population.

patients with HCC after surgery using a DL algorithm based on preoperative multi-parameter MRI were still limited. Additionally, most machine learning-based prediction models are not externally validated, and their robustness and capacity to generalize are yet to be verified.

This study aimed to assess the predictive value of a three-dimensional (3D) CNN model based on preoperative multi-parameter MRI for the OS of patients with HCC after surgery and compare it with that of clinical factors.

2. Materials and methods

2.1. Patients

Considering the retrospective nature of this study, the Institutional Review Board (B2021-214-01) of Sun Yat-sen University Cancer Center (SYSUCC) waived the requirement for informed consent. Patients who underwent GD-DTPA-enhanced MRI were enrolled in this study. The data of patients with HCC confirmed by surgery and pathology at SYSUCC between 2010 and 2015 and at Hunan Cancer Hospital between 2014 and 2016 were retrospectively collected. The inclusion criteria were as follows: (1) HCC confirmed via curative liver resection and postoperative pathological examination; (2) Child-Pugh class A or B; (3) required contrast-enhanced MRI scans within a month of surgery; and (4) complete information on follow-up and clinical data. Exclusion criteria included: (1) other malignancies; (2) unqualified artifacts on images; (3) extrahepatic metastases; (4) complications or rupture of HCC; and (5) other antitumor therapies performed before surgical resection. We included only the largest tumor if a patient had multiple HCCs. A total of 469 patients (407 men and 62 women, mean age 51.3 ± 11.3 years, range 30–86 years) with HCC from SYSUCC were included and randomly divided into the training (n = 326) and testing validation (n = 143) cohorts at a ratio of 7:3. We also enrolled 95 patients with HCC from Hunan Cancer Hospital (84 men and 11 women; mean age: 52.3 ± 9.5 years, range: 31–73 years) as the external validation cohort (shown in Fig. 1).

Table 1
Clinical characteristics of patients in the different cohorts.

Characteristic	Training cohort (N = 326)	Testing cohort (N = 143)	*P*1-value	External validation cohort (N = 95)	*P*2-value
BCLC stage			0.931		0.046
0	32	12		2	
A	231	105		79	
B	31	12		8	
C	32	14		6	
Sex			0.477		0.642
Male	280	127		84	
Female	46	16		11	
Age (Mean ± SD)	51.5 ± 11.5	50.7 ± 10.9	0.383	52.3 ± 9.5	0.817
HBsAg			0.026		0.903
Positive	289	115		84	
Negative	37	28		11	
Albumin (g/L)			0.593		<0.001
≥35	319	138		85	
< 35	7	5		10	
Total bilirubin (umol/L)			0.235		0.794
≤34	320	143		93	
> 34	6	0		2	
Prothrombin time (sec.)			0.37		<0.001
≤13.5	317	136		47	
> 13.5	9	7		48	
Cirrhosis			0.552		<0.001
Yes	225	94		64	
No	101	49		31	
AFP (ng/ml)			0.116		0.07
≤200	209	80		71	
> 200	117	63		24	
Child-Pugh class			0.317		<0.001
A	321	143		85	
B	5	0		10	
Tumor number			0.539		0.223
1	269	122		84	
≥2	57	21		11	
Maximum diameter(mm)	51.5 ± 31.5	48.4 ± 28.5	0.459	59.0 ± 29.0	0.007
Vascular invasion			0.94		0.274
Yes	32	14		6	
No	294	129		89	
Follow-up Median(IQR)	67(44–88)	63(36.5–77)	0.074	60(42–63)	<0.001

BCLC, Barcelona Clinic Liver Cancer; AFP, alpha-fetoprotein; Mean ± SD, mean ± standard deviation; Median (IQR), median (interquartile range). *P*1-value: training cohort vs testing cohort; *P*2-value: training cohort vs extral validation cohort.

The baseline clinical characteristics associated with postoperative prognosis, such as age, sex, history of hepatitis B, liver cirrhosis, serum AFP, albumin (ALB), total bilirubin, prothrombin time, Child-Pugh class (A or B), maximal tumor diameter, tumor number, vascular invasion, and BCLC stage, were collected (Table 1).

2.2. Follow-up

OS was the study endpoint. For the patients' medical records, contrast-enhanced CT or MRI was performed, and serum AFP concentrations were tested every 3 or 6 months to assess survival after surgery. The duration of survival of patients without post-operative medical records was sourced from their family members. The data was censored on December 30, 2021. OS was defined as the duration between HCC surgery and death, or the last follow-up.

2.3. MRI

MRI was performed at our hospital using a 3.0-T MRI system (Siemens Medical Solutions; GE Medical Systems; Philips Medical Systems) with a 16-channel phased-array coil within 1 month before the surgery. Our multi-parameter liver MRI sequences followed the standard protocol, including axial fat-suppressed T2-weighted imaging, axial in-phase and opposed-phase T1-weighted imaging (T1WI), diffusion-weighted imaging (DWI; b values of 0 and 800 s/mm^2), and liver acceleration volume acquisition-enhanced scanning (precontrast T1WI, T1WIpre; late arterial phase, AP; portal venous phase, PVP; delayed phase, DP). Dynamic contrast-enhanced liver acceleration volume acquisition images were acquired at 30–35s (AP), 50–55s (PVP), and 180s (DP) after extracellular contrast agent injection (gadolinium-diethylene triamine pentaacetate, GD-DTPA, 0.1 mmol/kg body weight, at a flow rate of 2.0 ml/s). Additionally, MRI was performed at Hunan Cancer Hospital using a Philips Achieca 3.0-T MRI system. The scanning parameters of the main sequences are shown in Table S1.

2.4. Tumor segmentation

Dynamic enhanced images are very important sequences in the diagnosis of HCC. Due to the image quality of T2WI and the lack of DWI and delayed phase sequences for a part of HCC patients, the volumes of interest (VOIs) of the tumors were manually delineated along the edge of the lesion on each transverse section from the three-phase images (T1WIpre, late AP, and PVP) covering the entire tumor by two independent radiologists with 4 and 5 years of abdominal imaging diagnostic experience using ITK-SNAP software (version 2.2.0; www.itksnap.org). All VOIs were drawn on slices with visible lesions (>10 mm). To ensure the reliability of the delineation, the physician was blinded to the other clinical and pathological data, except the HCCs. A schematic diagram of the outlined lesions is shown in Fig. 2 A-D. A senior radiologist with >30 years of abdominal imaging diagnostic experience reviewed and corrected the VOIs for ambiguous cases.

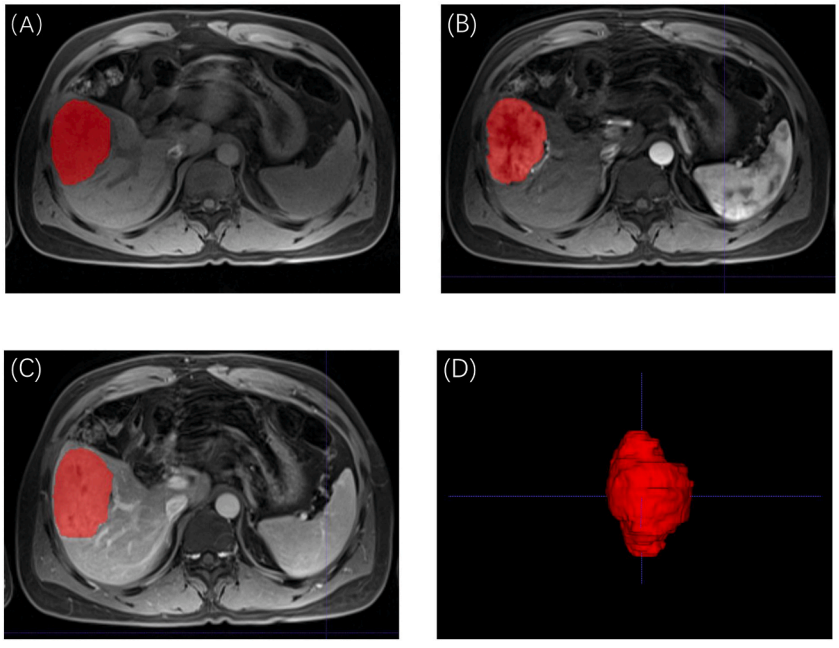

Fig. 2. Schematic diagram of the lesion delineation. (A–C) Axial T1WIpre, late arterial phase, and portal venous phase magnetic resonance images of hepatocellular carcinoma in the right liver lobe with the delineated volume of interest. (D) Three-dimensional structure of the lesion.

2.5. Image preprocessing and model construction

For spatial normalization, all images were resampled (to 1.0, 1.0, 1.0) and adjusted to a uniform orientation. All images of T1WIpre, AP, and PVP were normalized using signal intensity normalization, and all pixels outside the VOIs were set to zero. Since the original images were too large to be directly put into the network and the lesions only occupied a small amount of space, we centered on the lesions and cropped images into a uniform size of 100 × 100 × 100. The edge was enlarged by adding zero pixels and then cropped if the lesion was close to the edge. We generated a new four-dimensional array (3 × 100 × 100 × 100) and organized the images from the three modalities within this array, resulting in composite input images that effectively combined the data from all three modalities. The deep learning model was trained using classification labels assigned by a survival time threshold. For example, if the threshold was 5 years, the label was 1 when an event happened within 5 years; otherwise, the label was 0. The residual neural network (ResNet) developed by He et al. (2016) was chosen because it can reduce accuracy degradation in deeper networks. The 18-layer ResNet model (ResNet-18) used in our study (Fig. 3) consists of 17 convolutional layers, a fully connected layer, a max pooling layer, an average pooling layer, and a softmax layer for classification. Pre-processed images were input into the networks. We employed a 3 × 3 × 3 convolutional filter, doubling the number of filters when the feature map size was reduced by half. A convolutional layer with a stride of two was used to perform three rounds of downsampling. The extracted features were passed from the average pooling layer to the fully connected layer. A softmax layer produced the probability value of the event. We employed the focal loss as a loss function to tackle issues stemming from imbalanced data distribution. Two types of residual shortcut connections were inserted between the layers throughout the entire network. The first was used when the output and input had the same dimensions. The second connection was used when the dimension increased, and identity mapping with zero-filling at a stride of two was performed. Fig. 4 A-D illustrates the detailed model training process. For the training step, the batch size was 64, and the learning rate was 0.01. We conducted training over 100 epochs to achieve model convergence. To avoid overfitting and increase the robustness of the model, we employed the Python library of torchio (version: 0.18.88) to enlarge the datasets. Classic augmentation techniques, including flipping, translation, rotation, scaling, and adding gaussian noise, were used for each patient in our training cohort. During training, the AUC (area under the receiver operating characteristic curve) of the testing cohort of each epoch was recorded, and the final model was determined by the maximum AUC of testing. All experiments were conducted using an NVIDIA GeForce RTX 3090 workstation. We extracted the softmax layer's output to obtain the probability of each sample event's occurrence. We referred to this probability as the deep learning score (DL score), where a higher probability value signifies a greater likelihood of the event occurring within the sample. The link for the model code was offered via Github: https://github.com/WendyNice/HCC_multi_mod_survival.

2.6. Statistical analysis

Python (version 3.10.2) and R (version 4.0.4) were used for the statistical analysis. We employed independent-sample t-tests or Mann-Whitney U tests to compare continuous variables. These variables were presented as means (SD) for normally distributed data or medians (interquartile range [IQR]) for non-normally distributed data. Fisher's exact or chi-squared tests were used to compare categorical variables. We conducted backward variable selection to choose factors for constructing a multivariate clinical model based on the Akaike information criterion (AIC). This was achieved using the 'coxph' and 'AIC' function within the survival package in R. A combined Cox regression model was developed for the clinical factors and DL scores. Model performance was evaluated using the Concordance index (C-index), with values interpreted as follows: no predictive value (C-index = 0.5), low predictive value (0.5 < C-index <0.70), good predictive value (0.70 ≤ C-index <0.90), and excellent predictive value (C-index ≥0.90). The performances of the different models were compared using the cindex.p function of the SurvComp package in R. The Kaplan-Meier survival curves were compared using a log-rank test. All reported two-sided P-values <0.05 were considered significant, unless stated otherwise.

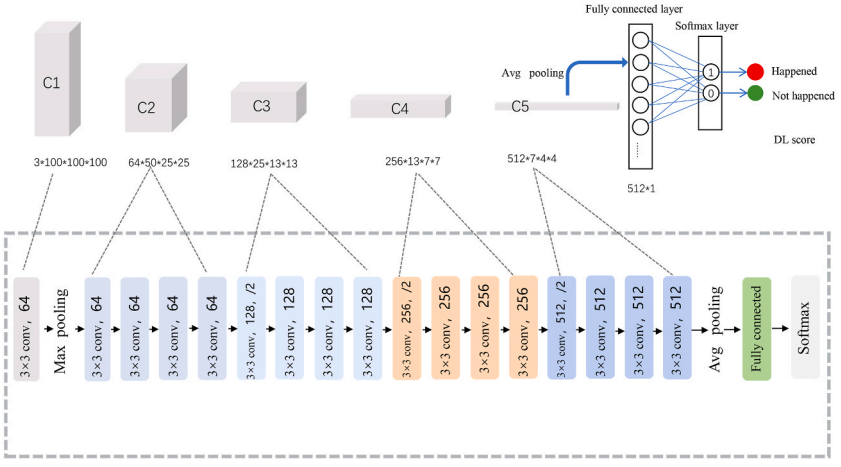

Fig. 3. ResNet-18 architecture.

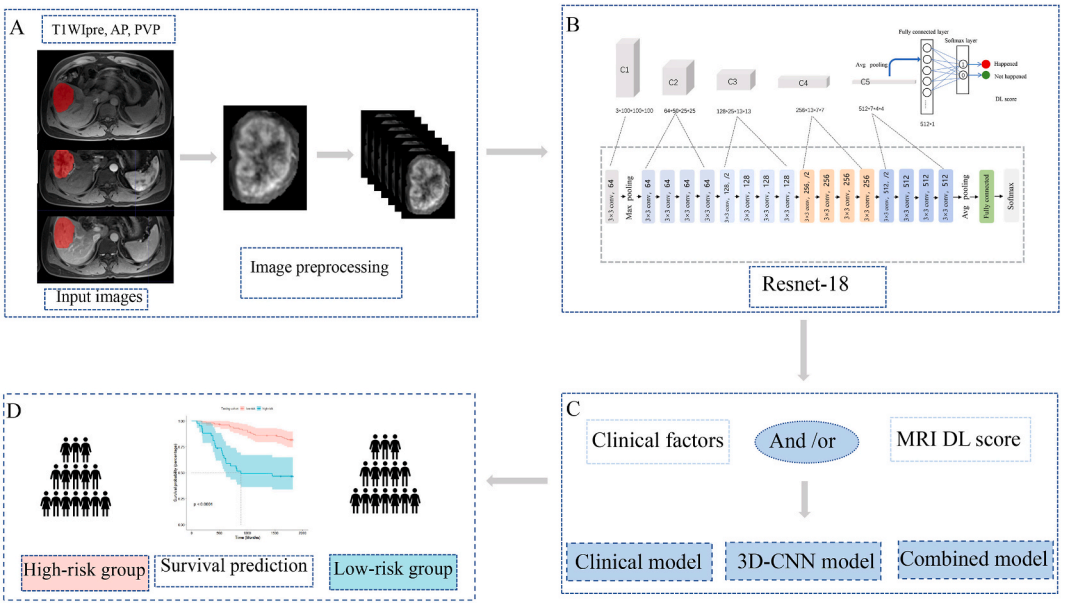

Fig. 4. Research layout. (A) Preprocessing of images (T1WIpre, late arterial phase, and portal venous phase), including translation, rotation, and zooming. (B) We trained the three-dimensional convolutional neural network (3D-CNN) model on three magnetic resonance imaging (MRI) sequences to extract features of the training cohort. A deep learning (DL) score outputted by the 3D-CNN model was assigned to each patient to predict overall survival. (C) Model construction: a combined model integrating the DL score and independent prognostic factors through Cox regression. (D) The cutoff value of the DL score was used to allocate patients into high- and low-risk groups.

3. Results

3.1. Clinical characteristics of the patients

A total of 564 patients (469 patients from SYSUCC and 95 patients from Hunan Cancer Hospital) were included in the analysis. We enrolled 44, 336, 43, and 46 patients with BCLC stages 0, A, B, and C from SYSUCC and 2, 79, 8, and 6 patients with BCLC stages 0, A, B, and C from Hunan Cancer Hospital, respectively. There were no significant differences in the clinical characteristics, except for HBsAg between the training and testing cohorts and the BCLC stage, ALB, PT, and cirrhosis between the training and external validation cohorts (Table 1). The median follow-up duration for the training, testing, and external validation cohorts was 67 (IQR, 44–88), 63 (IQR, 36.5–77), and 60 (IQR, 42–63) months, respectively (shown as a violin plot in Fig. S1). The mean 5-year OS for the training, testing, and external validation cohorts was 49 ± 19, 49 ± 19, and 50 ± 16 months, respectively, without statistical difference. The 5-year survival rates of the patients from Hunan Cancer Hospital and the external validation cohort were 73.5 % (345/469) and 65.3 % (62/95), respectively.

3.2. Clinical model development

Multivariate Cox regression analysis of the training cohort data according to AIC showed that the BCLC stage ((hazard ratios, HR) = 2.240; 95 % CI: 1.732–2.901; P < 0.01) and total ALB (HR = 3.554; 95 % CI: 1.224–10.318; P = 0.019) were independent predictors of postoperative OS and could be used to construct the clinical model (Table 2). The C-index of the clinical model (AIC = 914.10) in

Table 2
Multivariable Cox regression analysis of the clinical factors associated with the overall survival.

Variable	Multivariate analysis		
	HR	95 % CI	*P* value
BCLC stage	2.240	(1.732, 2.901)	<0.01
Sex	1.741	(0.953, 3.188)	0.071
Age	0.995	(0.976, 1.010)	0.607
Albumin (g/L)	3.554	(1.224, 10.318)	0.019
Total bilirubin(umol/L)	0.15	(0.019, 1.257)	0.081
AFP	1.20	(0.757, 1.911)	0.434

CI, confidence interval; AFP, alpha-fetoprotein; BCLC, Barcelona Clinic Liver Cancer; HR, hazard ratio.
Selecting significant features according to Akaike information criterion (AIC).

predicting 5-year OS was 0.675 (95 % CI: 0.627–0.723), 0.674 (95 % CI: 0.604–0.744), and 0.631 (95 % CI: 0.558–0.704) for training, testing, and external validation cohorts, respectively.

3.3. Performance of the deep learning model

The performance of the deep learning model based on the DL score from the model trained by T1WIpre, late AP, and PVP sequence features was evaluated using cox regression. The HR of the DL score was 51.491 (95 % CI: 19.080–139), which was analyzed to predict the OS of patients with HCC. The C-index of the 3D-CNN model (AIC = 892.39) in prediting 5-year OS was 0.746 (95 % CI: 0.685–0.810), 0.714 (95 % CI: 0.629–0.804), and 0.698 (95 % CI: 0.597–0.800) for training, testing, and external validation cohorts, respectively.

The performance of the combined model integrating the DL score of the three sequences and clinical features (BCLC stage and total ALB) was slightly improved. The C-index of the combined model (AIC = 881.93) in predicting 5-year OS was 0.753 (95 % CI: 0.691–0.816), 0.750 (95 % CI: 0.659–0.841), and 0.723 (95 % CI: 0.623–0.823) for the training, testing, and external validation cohorts, respectively. We calculated the cutoff value of the DL score to distinguish the patients with high (DL score >0.55) and low (DL score ≤0.55) mortality risks using X-tile software. The Kaplan-Meier survival curves for the OS of the cohorts determined by the three different models are shown in Fig. 5 (P < 0.001, 0.036, and 0.350 for the clinical model [Fig. 5 D-F] and P all < 0.001 for the 3D-CNN model [Fig. 5 A-C] and combined model [Fig. 5 G-I] in the training, testing, and external validation cohorts, respectively).

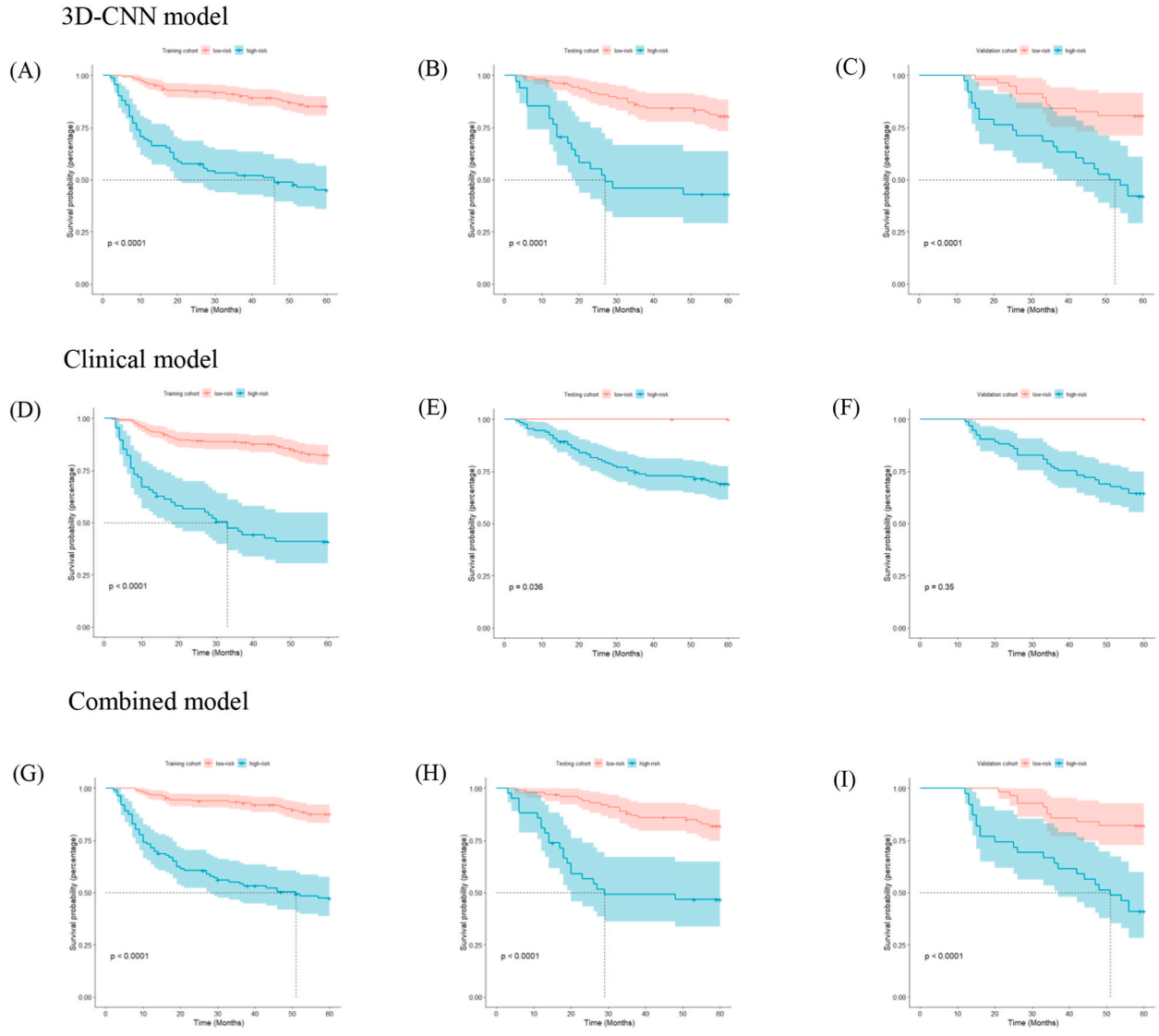

Fig. 5. Kaplan-Meier curves for overall survival for high- and low-risk groups based on different models. (A–C) The 3D-CNN model. (D–F) The clinical model. (G–I) The combined model. 3D-CNN, three-dimensional convolutional neural network.

Table 3
Comparison of model performance in predicting the overall survival.

Cohort	Time(year)	Clinical model AIC	Clinical model_ C-index (95 % CI)	3D-CNN model AIC	3D-CNN model C-index (95 % CI)	Combined model AIC	Combined model C-index (95 % CI)	P1 value	P2 value	P3 value
Training	1	421.540	0.734(0.662,0.807)	405.257	0.808(0.716,0.900)	398.230	0.820(0.728,0.911)	0.048	0.004	0.244
Testing	1	421.540	0.758(0.615,0.900)	405.257	0.731(0.550,0.912)	398.230	0.807(0.626,0.989)	0.609	0.241	0.025
External validation	1	421.540	0.394(0.000,0.830)	405.257	0.872(0.301,1.000)	398.230	0.809(0.237,1.000)	0.048	0.041	0.785
Training	3	711.990	0.709(0.653,0.764)	691.045	0.787(0.716,0.857)	678.158	0.798(0.727,0.868)	0.011	< 0.001	0.205
Testing	3	711.990	0.692(0.616,0.768)	691.045	0.717(0.619,0.816)	678.158	0.761(0.662,0.860)	0.311	0.036	0.019
External validation	3	711.990	0.604(0.513,0.695)	691.045	0.663(0.539,0.787)	678.158	0.686(0.562,0.810)	0.167	0.056	0.094
Training	5	914.099	0.676(0.627,0.724)	892.386	0.748(0.686,0.811)	881.929	0.754(0.692,0.817)	0.009	< 0.001	0.303
Testing	5	914.099	0.674(0.604,0.745)	892.386	0.715(0.624,0.806)	881.929	0.751(0.659,0.842)	0.204	0.017	0.029
External validation	5	914.099	0.632(0.558,0.705)	892.386	0.698(0.597,0.800)	881.929	0.723(0.622,0.825)	0.092	0.016	0.036

AIC, Akaike information criterion; 3D-CNN, three-dimensional convolutional neural network; C-index, concordance index; CI, confidence interval.
P1 value:3D-CNN model vs clinical model; P2 value: combined model vs clinical model; P3 value: 3D-CNN model vs combined model.

3.4. Performance comparison

The performance of models was compared by the C-index and Delong test. The C-index of the clinical model in predicting 5-year OS was significantly lower than those of the 3D-CNN and combined models for the training cohort ($P = 0.009$ and $P < 0.001$, respectively). For the testing and external validation cohorts, the C-index of the 3D-CNN model was slightly higher than that of the clinical model ($P = 0.204$ and 0.092, respectively). The combined model was significantly more accurate in predicting outcomes than the clinical model ($P = 0.017$ and 0.016, respectively) and the 3D-CNN model ($P = 0.029$ and 0.036, respectively) in the testing and external cohorts (Table 3). Besides, the C-index of three different models for predicting 1-, 3-, and 5-year OS is shown in Table 3.

4. Discussion

DL algorithms are gaining attention for image recognition, disease prediction, and prognostic prediction [16]. This study established and validated a combined model integrating clinical factors and DL features extracted from T1WIpre, AP, and PVP images for the preoperative prediction of postoperative OS. Our results showed that the 3D-CNN model had a better predictive value than the clinical model for the training and external validation cohorts. The combined model had significantly better predictive ability than the clinical or 3D-CNN model. Overall, the 3D-CNN model was useful for predicting OS after surgical resection.

Medical imaging has been enhanced with convolutional neural networks, such as LeNet-5, AlexNet, VGGNet, and Google Inception Net [16,25]. However, the performance of the networks above may decline with deeper neural network layers owing to gradient vanishing or diffusion and degeneration in the network model. ResNet was selected in this study as it can fully train the underlying network from cross-layer connections, overcome problems such as degradation and accuracy decline, and optimize the underlying parameters better than traditional non-residual-based DL methods [26].

Several studies have established different models based on DL algorithms or radiomics to predict prognostic indicators, such as survival and recurrence in patients with HCC after treatment using other images. Saillard et al. [24] used pathological images to predict the postoperative survival of patients with HCC based on a 2D-CNN model with a C-index of 0.75, similar to our study. However, pathological slides can only be obtained after surgery and reflect the local tumor heterogeneity rather than the whole tumor. Some authors used pre-stereotactic body radiotherapy CT images based on CNN to predict OS with a C-index of 0.580, which was significantly lower than that of our CNN model [27]. Wang et al. [15] built a clinical-radiomic model to predict the 5-year OS based on MRI images with a C-index of 0.75 in the validation cohort, similar to our study. Whereas, the C-index between training and validation cohorts was far different (0.98 vs. 0.75), indicating the possibility of overfitting. Liu et al. [28] developed a DL-radiomics model using preoperative contrast-enhanced ultrasound images for early-stage HCC to predict progression-free survival after ablation or surgical resection. And some scholars used the radiomics model based on single-center CT images to predict OS of HCC patients after surgery [29,30]. The performance of these models was similar to that of our study, while the generalization performance of the model needs to be further validated with limitations of the single-center study. Compared with the above studies, our cohort included multicenter patients with multi-parameter MRI, which not only can provide more high-throughput, high-dimensional microscopic information and high soft tissue resolution but also improved the accuracy and robustness of the model. Additionally, we used the whole volume of the tumor to build the model to take full advantage of the tumor information. Moreover, compared with the abovementioned radiomics models, the CNN model only needs minor data preprocessing, which then extracts informative representations in a self-learning manner. The C-index of the 3D-CNN model was >0.7 for the training, testing, and external validation cohorts with good robustness. The prediction of high-risk and low-risk patients based on 3D-CNN or combined model showed different prognosis in three different cohorts with statistical significance ($P < 0.001$). It will be helpful to relieve the financial burden of the low-risk patients and shorten their waiting time for further treatment to reduce the mortality risk of the high-risk patients.

Our study suggests that the preoperative BCLC stage and ALB are independent clinical predictors for OS after HCC resection based on multivariate Cox regression. The BCLC staging system [31] considers tumor burden, which covers an evaluation of the size and number of the tumors, as well as the presence or absence of vascular invasion, extrahepatic metastasis, liver function, and systemic status. Later stages of HCC are associated with worse prognoses. The ALB concentrations reflect liver function, and HCC's prognosis is closely linked to liver function and stratified by the albumin-bilirubin balance for compensated liver function [32]. However, the C-index of the clinical model for predicting OS was 0.675, 0.674, and 0.631 for the training, testing, and external validation cohorts, respectively; the 3D-Resnet CNN model had a better performance. As 3D-CNNs can identify complex features and extract information from MR images in self-supervised or unsupervised learning, they can achieve automatic end-to-end survival prediction. Combined models performed best, which shows that integrating MRI and clinical features can help complement each other to improve predictions and reveal extreme variability.

5. Limitations

There are some limitations to our study. First, it was a retrospective study that had some selection bias. Second, this study only included OS as the endpoint; other survival indicators, such as PFS, should be included in further research. Third, we only used T1WIpre, AP, and PVP sequences in this study. Due to the image quality of T2WI, when we added the T2WI sequence to the model for training, the performance of the model didn't increase but slightly decreased. Besides, a part of the patients lacked the DWI and delayed phase sequences. We should further study the values of the above sequences in the prediction of HCC prognosis using the DL method. Besides, we only included preoperative characteristics to build models; some postoperative features, such as adjuvant treatments, should be studied in our further research.

6. Conclusion

In conclusion, deep learning algorithms based on multi-parameter contrast-enhanced MRI can preoperatively predict the OS of patients with HCC with good accuracy. Moreover, the combined model integrating the DL score and clinical factors showed a better predictive value than the clinical or 3D-CNN models and may be a better option.

Ethics statement

This study was reviewed and approved by Institutional Review Board of Sun Yat-sen University Cancer Center, with the approval number B2021-214-01. Informed consent was not required because this was a retrospective study.

Data availability statement

Data will be made available from the corresponding author, Chuanmiao Xie, upon reasonable request.

CRediT authorship contribution statement

Lidi Ma: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Congrui Li:** Validation, Formal analysis, Data curation. **Haixia Li:** Validation, Methodology, Formal analysis. **Cheng Zhang:** Data curation. **Kan Deng:** Methodology, Formal analysis. **Weijing Zhang:** Funding acquisition. **Chuanmiao Xie:** Resources, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:Haixia Li reports a relationship with Bayer Corporation that includes: employment. Kan Deng reports a relationship with Philips Healthcare that includes: employment. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e31451.

References

[1] H. Sung, J. Ferlay, R.L. Siegel, et al., Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA A Cancer J. Clin. 71 (2021) 209–249.
[2] Hepatocellular carcinoma, Nat. Rev. Dis. Prim. 7 (2021) 7.
[3] K. Ito, N. Takemura, F. Inagaki, F. Mihara, N. Kokudo, Difference in treatment algorithms for hepatocellular carcinoma between world's principal guidelines, Global health & medicine 2 (2020) 282–291.
[4] J. Marrero, L. Kulik, C. Sirlin, et al., Diagnosis, staging, and management of hepatocellular carcinoma: 2018 practice guidance by the American association for the study of liver diseases, Hepatology 68 (2018) 723–750.
[5] J. Yang, P. Hainaut, G. Gores, A. Amadou, A. Plymoth, L. Roberts, A global view of hepatocellular carcinoma: trends, risk, prevention and management, Nat. Rev. Gastroenterol. Hepatol. 16 (2019) 589–604.
[6] S. Shimada, T. Kamiyama, T. Kakisaka, et al., Impact of gadolinium-ethoxybenzyl-diethylenetriamine pentaacetic acid-enhanced magnetic resonance imaging on the prognosis of hepatocellular carcinoma after surgery, JGH open 5 (2021) 41–49.
[7] L. Zhang, P. Luo, L. Chen, et al., Model to predict overall survival in patients with hepatocellular carcinoma after curative hepatectomy, Front. Oncol. 10 (2020) 537526.
[8] A. Chan, J. Zhong, S. Berhane, et al., Development of pre and post-operative models to predict early recurrence of hepatocellular carcinoma after surgical resection, J. Hepatol. 69 (2018) 1284–1293.
[9] L. Ma, K. Deng, C. Zhang, et al., Nomograms for predicting hepatocellular carcinoma recurrence and overall postoperative patient survival, Front. Oncol. 12 (2022) 843589.
[10] J. Zeng, J. Zeng, J. Liu, J. Zeng, Development of pre and post-operative nomograms to predict individual survival for ideal liver resection candidates with hepatocellular carcinoma, Liver Int. 41 (2021) 2974–2985.
[11] R. Liao, X. Wei, P. Che, K. Yin, L. Liu, Nomograms incorporating the CNLC staging system predict the outcome of hepatocellular carcinoma after curative resection, Front. Oncol. 11 (2021) 755920.
[12] L. Brenet Defour, S. Mulé, A. Tenenhaus, et al., Hepatocellular carcinoma: CT texture analysis as a predictor of survival after surgical resection, Eur. Radiol. 29 (2019) 1231–1239.
[13] J. Heimbach, L. Kulik, R. Finn, et al., AASLD guidelines for the treatment of hepatocellular carcinoma, Hepatology 67 (2018) 358–380.
[14] EASL clinical practice guidelines: management of hepatocellular carcinoma, J. Hepatol. 69 (2018) 182–236.
[15] X. Wang, L. Long, Y. Cui, et al., MRI-based radiomics model for preoperative prediction of 5-year survival in patients with hepatocellular carcinoma, Br. J. Cancer 122 (2020) 978–985.
[16] G. Litjens, T. Kooi, B. Bejnordi, et al., A survey on deep learning in medical image analysis, Med. Image Anal. 42 (2017) 60–88.
[17] E. Montagnon, M. Cerny, A. Cadrin-Chênevert, et al., Deep learning workflow in radiology: a primer. Insights into Imaging 11:22, 2020.
[18] D. Ardila, A. Kiraly, S. Bharadwaj, et al., End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography, Nat. Med. 25 (2019) 954–961.

[19] S. Wang, J. Shi, Z. Ye, et al., Predicting EGFR mutation status in lung adenocarcinoma on computed tomography image using deep learning, Eur. Respir. J. 53 (2019).
[20] S. Wang, X. Han, J. Du, et al., Saliency-based 3D convolutional neural network for categorising common focal liver lesions on multisequence MRI, Insights into imaging 12 (2021) 173.
[21] K. Yasaka, H. Akai, A. Kunimatsu, O. Abe, S. Kiryu, Liver fibrosis: deep convolutional neural network for staging by using gadoxetic acid-enhanced hepatobiliary phase MR images, Radiology 287 (2018) 146–155.
[22] K. Yasaka, H. Akai, O. Abe, S. Kiryu, Deep learning with convolutional neural network for differentiation of liver masses at dynamic contrast-enhanced CT: a preliminary study, Radiology 286 (2018) 887–896.
[23] Y. Zhang, X. Lv, J. Qiu, et al., Deep learning with 3D convolutional neural network for noninvasive prediction of microvascular invasion in hepatocellular carcinoma, J. Magn. Reson. Imag. 54 (2021) 134–143.
[24] C. Saillard, B. Schmauch, O. Laifa, et al., Predicting survival after hepatocellular carcinoma resection using deep learning on histological slides, Hepatology 72 (2020) 2000–2013.
[25] M. Mazurowski, M. Buda, A. Saha, M. Bashir, Deep learning in radiology: an overview of the concepts and a survey of the state of the art with focus on MRI, J. Magn. Reson. Imag. 49 (2019) 939–954.
[26] K. He, X. Zhang, S. Ren, S. Jian, Deep residual learning for image recognition, in: IEEE Conference on Computer Vision & Pattern Recognition, 2016, https://doi.org/10.1109/CVPR.2016.90.
[27] L. Wei, D. Owen, B. Rosen, et al., A deep survival interpretable radiomics model of hepatocellular carcinoma patients, Phys. Med. 82 (2021) 295–305.
[28] F. Liu, D. Liu, K. Wang, et al., Deep learning radiomics based on contrast-enhanced ultrasound might optimize curative treatments for very-early or early-stage hepatocellular carcinoma patients, Liver Cancer 9 (2020) 397–413.
[29] Q. Liu, J. Li, F. Liu, et al., A radiomics nomogram for the prediction of overall survival in patients with hepatocellular carcinoma after hepatectomy, Cancer Imag. 20 (2020) 82.
[30] P. Deng, B. Zhao, X. Huang, et al., Preoperative contrast-enhanced computed tomography-based radiomics model for overall survival prediction in hepatocellular carcinoma, World J. Gastroenterol. 28 (2022) 4376–4389.
[31] M. Reig, A. Forner, J. Rimola, et al., BCLC strategy for prognosis prediction and treatment recommendation: the 2022 update, J. Hepatol. 76 (2022) 681–693.
[32] P. Johnson, S. Berhane, C. Kagebayashi, et al., Assessment of liver function in patients with hepatocellular carcinoma: a new evidence-based approach-the ALBI grade, J. Clin. Oncol. 33 (2015) 550–558.