A Deep Learning Nomogram for the Prediction of Early Recurrence in Hepatocellular Carcinoma After Curative Surgery

Meng Yan  
Jinan University First Affiliated Hospital

Xiao Zhang  
Zhuhai City People's Hospital

Bin Zhang  
Jinan University First Affiliated Hospital

Zhijun Geng  
Sun Yat-sen University Cancer Center

Chuanmiao Xie  
Sun Yat-sen University Cancer Center

Wei Yang  
Southern Medical University

Shuixing Zhang  
Jinan University First Affiliated Hospital

Zhendong Qi  
Zhuijiang Hospital

Ting Lin  
Zhuijiang Hospital

Qiying Ke  
GZHUCM: Guangzhou University of Chinese Medicine

Xinming Li  
Zhuijiang Hospital

Shutong Wang  
Sun Yat-sen University First Affiliated Hospital

Xianyue Quan  
quanxianyue2014@163.com  
Zhuijiang Hospital  
https://orcid.org/0000-0003-2293-9345

Research Article

Keywords: Liver Cancer, Artificial Intelligence, Convolutional neural network, Machine Learning, Radiomics, Magnetic Resonance Imaging, Gd-EOB-DTPA, Prognosis, Prediction Model, Risk factors
Abstract

Purpose: The accurate prediction of post-hepatectomy early recurrence in patients with hepatocellular carcinoma (HCC) is crucial for decision-making regarding postoperative adjuvant treatment and monitoring. We aimed to develop and validate a deep-learning (DL) nomogram based on MRI for predicting early recurrence in HCC after curative resection.

Methods: We retrospectively included 285 HCC patients who underwent Gadolinium-ethoxybenzyl-diethylenetriamine pentaacetic acid (Gd-EOB-DTPA)-enhanced MRI within one month before curative resection. Deep features were extracted from images of the arterial phase (AP), portal venous phase (PVP), and hepatobiliary phase (HBP) using VGGNet-19. Pearson's correlation was firstly used to exclude redundant features. Three feature selection methods and five classification methods were combined to construct the DL signature. Univariate and multivariate logistic regression analyses were used to identify the independent risk factors for the early recurrence, which were incorporated into the DL nomogram.

Results: Microvascular invasion (P = 0.039), tumor number (P = 0.001), and three-phase-based DL signature (P<0.0001) were independent risk factors for early recurrence. The DL nomogram integrating the DL signature and clinical risk factors outperformed the clinical nomogram which combined clinical risk factors, in the training set (AUC: 0.949 vs. 0.751; P<0.0001) and validation set (AUC: 0.908 vs. 0.712; P = 0.002). Excellent calibration was achieved for the DL nomogram in both training and validation sets. Decision curve analysis confirmed the clinical usefulness of the DL nomogram.

Conclusions: The proposed DL nomogram was superior to the traditional clinical nomogram in predicting early recurrence for HCC patients after curative resection.

Introduction

Hepatic resection is the first-line treatment for patients with early-stage Hepatocellular carcinoma (HCC) and well-preserved liver function [1]. However, the recurrence rate of HCC is as high as 70% five years after surgery Ishizawa et al. [2]. More than 80% of recurrences are intrahepatic, including intrahepatic metastases from a preliminary tumor, considered as true recurrence, and de novo multicentric metastasis [3]. The poor prognosis of patients is closely related to intrahepatic metastases and mainly presents as early recurrence (within 2 years), whereas late recurrence (>2 years) is more likely to be associated with underlying liver diseases, such as cirrhosis [4]. Thus, the assessment of the risk of early recurrence in HCC patients is clinically relevant.

Numerous cancer-related factors have been identified as predictors of early recurrence, such as microvascular invasion (MVI), surgical margin, and tumor size [5, 6]. However, most of the risk factors can only be obtained by postoperative pathology and cannot be used to assess prognosis and develop treatment plans prior to hepatectomy. Medical imaging is a routine preoperative examination for patients with HCC. Gadolinium-ethoxybenzyl-diethylenetriamine pentaacetic acid (Gd-EOB-DTPA) is a
hepatobiliary-specific contrast agent, and Gd-EOB-DTPA-enhanced magnetic resonance imaging (MRI) is better than [7] MRI using other agents [7].

Although several imaging features observed on Gd-EOB-DTPA-enhanced MRI, including arterial peritumoral enhancement, non-smooth tumor margin, and peritumoral hypointensity on hepatobiliary phase (HBP) [8], have been demonstrated to be associated with early recurrence in HCC patients, the prediction of early recurrence by MR features may be subjective and dependent on the radiologist’s experience. Visual features are limited by image grayscale and lose much information about tumor heterogeneity.

Deep learning (DL), a subset of machine learning, is a new diagnostic technology for mining the internal information of medical images. DL can be applied to prognosis prediction [9], and treatment response evaluation [10] by automatically extracting deep-learned or high-order image features. Among them, convolutional neural network (CNN) is famous for handling image classification tasks [11]. Thus, if DL-based image features can be directly used to predict prognosis, it would provide a promising non-invasive method to provide better options for individualized patient treatment.

Therefore, we aimed to investigate the feasibility of deep features extracted from Gd-EOB-DTPA-enhanced MR images for predicting the early recurrence of HCC after curative resection. Furthermore, we evaluated the predictive performance of the DL-based nomogram incorporating deep features and significant clinical indicators.

Materials And Methods

Patients and dataset

The institutional ethics review boards of Zhujiang Hospital of Southern Medical University (2020-KY-094-01) and Sun Yat-Sen University Cancer Center (SL-B2021-214-02) approved this retrospective study, and the requirement for informed consent was waived. Patients with suspected HCC who underwent Gd-EOB-DTPA-enhanced MRI scans between January 2012 and September 2018 at two institutions prior to curative resection were consecutively included. The inclusion criteria were as follows: (a) patients with pathological confirmation of HCC; (b) patients with Barcelona Clinic Liver Cancer (BCLC) stage 0, A, or B HCC; (c) patients received no previous anti-cancer treatment; and (d) patients who underwent Gd-EOB-DTPA-enhanced MRI of the liver within one month before surgery. The exclusion criteria were as follows: (a) recurrent HCC or combined with hepatocyte cholangiocarcinoma or metastatic tumor in the liver; (b) without radiographic MVI or extrahepatic metastasis; (c) incomplete clinical, radiological, pathological, or follow-up data; and (d) patients died due to postoperative complications or liver cancer rupture within two weeks (Supplementary Fig. 1). All patients were randomly divided into the training and validation sets at a ratio of 7:3.

Baseline clinicopathological data were collected from electronic medical records. Clinical data included demographics and time to early recurrence. Laboratory features included neutrophil count, Hepatitis B
virus DNA, α-fetoprotein (AFP) level, alanine aminotransferase (ALT), aspartate aminotransferase (AST), and γ-Glutamyltranspeptidase (GGT). Pathologic data were the presence of MVI, defined as tumor emboli in a vascular space lined by endothelial cells on microscopy [12], Barcelona Clinic Liver Cancer (BCLC) stage, and tumor number.

**Follow-up surveillance and clinical endpoint**

All patients were followed up for at least two years after curative resection. Patients were screened for tumor recurrence through serum AFP level, ultrasound, contrast CT or MRI scan of the chest and abdomen in the first month after surgery and once every three months thereafter during the first year and every six months thereafter. The censored follow-up time was October 1, 2020.

The study endpoint was early recurrence, which was defined as one (or more) of the following events that occurred within two years after curative resection: (a) presence of new hepatic lesions with typical imaging findings of HCC, (b) atypical imaging findings with biopsy or re-postoperative pathology-confirmed HCC, or postoperative transarterial chemoembolization-revealed tumor staining, and (c) extrahepatic metastases.

**MR imaging acquisition**

The MRI machines and parameters are provided in *Supplementary Methods* and *Supplementary Table 1*.

**Qualitative analysis of MR Images**

Two radiologists with 5 and 10 years of experience in diagnostic abdominal imaging independently observed the imaging features without prior knowledge of the pathological findings, and when they disagreed on the results, the decision was made by mutual agreement. The MR features included: (a) tumor size, defined as maximum diameter on transverse HBP images; (b) arterial phase (AP) enhancement type: type 1 represents a homogeneous enhancement pattern with no increased arterial blood flow; type 2 represents a homogeneous enhancement with increased arterial blood flow; type 3 represents a heterogeneous enhancement containing non-enhanced areas; type 4 represents a heterogeneous enhancement pattern with irregular ring-like structures [13-15]; type 5 represents a heterogeneous and hypointense; (c) capsule appearance: peripheral rim of uniform and smooth hyperenhancement in the portal phase (PVP) or delayed phase, and categorized into three groups (absent, incomplete and complete) [16]; (d) hypodense halo: a rim of hypointensity partially or wholly surrounding the tumor; (e) intratumor necrosis, a low signal on T1-weighted imaging (T1WI), a high signal on T2-weighted imaging (T2WI), and a low signal on all enhanced phases; (f) satellite nodules, defined as small (< 2 cm) tumor nodules close (< 2 cm) to the main tumor [17]; (g) peritumoral hypointense, defined as a flame-like or wedge-shaped hypointense areas of the hepatic parenchyma around tumor on HBP images [18]. *Supplementary Fig. 2* shows the images of the MR features.

**Image segmentation and DL feature extraction**
The regions of interest were delineated around the boundary of tumor at the largest dimension. A state-of-the-art architecture VGGNet-19 was then applied to extract 1472 DL features from the AP, PVP, and HBP images, respectively. The DL network contains five convolutional layers, four max-pooling layers, three fully connected layers, and a softmax layer. The DL workflow is shown in Fig. 1. More details are provided in Supplementary Materials.

**Feature selection and deep learning signature development**

The DL features were subjected to the following steps: feature value preconditioning, de-redundancy, and dimensionality reduction to select the features strongly related to early recurrence; then, machine learning methods were used to predict the status of outcome events and establish a DL signature that can predict early recurrence. All features were first normalized to the range of [0,1] by the minimum-maximum normalization method. Moreover, Spearman correlation analysis was added to retain features associated with the early recurrence of HCC (P < 0.05). Then, the Pearson correlation coefficient (i.e., r) was used to remove one redundant feature with a lower r from the feature pairs (r > 0.9). The high predictive features obtained were further screened by variance analysis, recursive feature elimination (RFE), and Relief algorithm. Five types of classifiers, including random forest (RF), support vector machine (SVM), least absolute shrinkage and selection operator logistic regression (LASSO), Adaboost, and Gaussian Process (GP), were compared to identify the outcome status of early recurrence for every phase of the DL features.

**Clinical and deep learning analysis**

A two-sided P-value was considered statistically significant if < 0.10. Univariate logistic regression analysis was performed in the training set, and significant variables P < 0.10 were entered into the multivariate logistic regression using the forward likelihood ratio method to identify the independent risk factors for early recurrence. A two-sided P-value < 0.05 was considered statistically significant. The nomogram was plotted based on the results of multivariate logistic regression models.

Collinearity analysis of conventional clinical factors and DL signatures was also performed. The evaluation indicators were tolerance and variance inflation factor (VIF), tolerance value < 0.1, or a VIF value > 5, considered as collinearity between two variables.

**Statistical analysis**

Comparisons between two groups were conducted using the Chi-square test or Fisher's exact test for categorical variables while the Mann-Whitney U test for continuous variables.

The receiver operating characteristic curve (ROC) analysis was employed to calculate the area under the curve (AUC), accuracy, sensitivity, and specificity. Comparisons between different DL signatures, and between different models were performed using the Delong test. Model fit was assessed using calibration plots using 1000 bootstrap resamples. The clinical utility of the models was evaluated using decision
curve analysis (DCA). Softwares and packages for statistical analyses are provided in Supplementary Materials. All statistical tests were two-sided with a significance level of 0.05.

Results

Clinical characteristics

A total of 285 patients (male, n = 254; mean age, 52.89 years ± 11.69, range, 13-79 years) were included, which was divided into the training set (n = 195) and validation set (n = 90), respectively. Early recurrence occurred in 77 (27.0%) patients and there was no difference in the early recurrence rate between the training set (27.7%, 54/195) and validation set (25.6%, 23/90). The clinical characteristics of the two groups are shown in Table 1. No statistically significant difference was observed between the two sets (P = 0.373-1.000).
Table 1
Baseline patients characteristics in the training and validation sets

| Characteristics     | Total (n = 285) | Training set (n = 195) | Validation set (n = 90) | P-value |
|---------------------|----------------|------------------------|-------------------------|---------|
| Age (years)         | 54.0 (45.0-61.0) | 52.0 (45.0-60.0)       | 56.5 (45.0-62.0)        | 0.113   |
| Gender              |                |                        |                         | 0.153   |
| Male                | 254 (89.1)     | 170 (87.2)             | 84 (93.3)               |         |
| Female              | 31 (10.9)      | 25 (12.8)              | 6 (6.7)                 |         |
| NE (10⁹/L)          | 3.6 (2.8-4.6)  | 3.6 (2.9-4.6)          | 3.4 (2.8-4.7)           | 0.622   |
| HBV-DNA (IU/ml)     | 3.9×10³ (0.0-4.7×10⁵) | 3.1×10³ (0.0-3.7×10⁵) | 8.4×10³ (0.0-8.3×10⁵)   | 0.608   |
| AFP (ng/ml)         |                |                        |                         | 0.513   |
| ≤400                | 232 (81.4)     | 161 (82.6)             | 71 (78.9)               |         |
| >400                | 53 (18.6)      | 34 (17.4)              | 19 (21.1)               |         |
| ALT (U/L)           |                |                        |                         | 0.670   |
| ≤50                 | 206 (72.3)     | 139 (71.3)             | 67 (74.4)               |         |
| >50                 | 79 (27.7)      | 56 (28.7)              | 23 (25.6)               |         |
| AST (U/L)           |                |                        |                         | 0.484   |
| ≤40                 | 204 (71.6)     | 137 (70.3)             | 67 (74.4)               |         |
| >40                 | 81 (28.4)      | 58 (29.7)              | 23 (25.6)               |         |
| GGT (U/L)           |                |                        |                         | 0.600   |
| ≤60                 | 176 (61.8)     | 118 (60.5)             | 58 (64.4)               |         |
| >60                 | 109 (38.2)     | 77 (39.5)              | 32 (35.6)               |         |
| BCLC stage          |                |                        |                         | 0.757   |
| 0                   | 61 (21.4)      | 43 (22.1)              | 18 (20.0)               |         |
| A/B                 | 224 (79.4)     | 152 (77.9)             | 72 (80.0)               |         |
| Tumor size (mm)     |                |                        |                         | 0.391   |
| ≤30                 | 150 (52.6)     | 106 (54.4)             | 44 (48.9)               |         |
| Characteristics                  | Total (n = 285) | Training set (n = 195) | Validation set (n = 90) | P-value |
|---------------------------------|-----------------|------------------------|-------------------------|---------|
| >30                             | 135 (47.4)      | 89 (45.6)              | 46 (51.1)               |         |
| **AP enhancement type**         |                 |                        |                         | 0.960   |
| Type 1                          | 11 (3.9)        | 8 (4.1)                | 3 (3.3)                 |         |
| Type 2                          | 62 (21.8)       | 43 (22.1)              | 19 (21.1)               |         |
| Type 3                          | 168 (58.9)      | 112 (57.4)             | 56 (62.2)               |         |
| Type 4                          | 20 (7.0)        | 15 (7.7)               | 5 (5.6)                 |         |
| Type 5                          | 24 (8.4)        | 17 (8.7)               | 7 (7.8)                 |         |
| **Capsule appearance**          |                 |                        |                         | 1.000   |
| Absent                          | 101 (35.4)      | 69 (35.4)              | 32 (35.6)               |         |
| Incomplete                      | 69 (24.2)       | 47 (24.1)              | 22 (24.4)               |         |
| Complete                        | 115 (40.4)      | 79 (40.5)              | 36 (40)                 |         |
| **Hypodense halo**              |                 |                        |                         | 0.383   |
| Absent                          | 259 (90.9)      | 175 (89.7)             | 84 (93.3)               |         |
| Present                         | 26 (9.1)        | 20 (10.3)              | 6 (6.7)                 |         |
| **Intratumor necrosis**         |                 |                        |                         | 0.402   |
| Absent                          | 201 (70.5)      | 141 (72.3)             | 60 (66.7)               |         |
| Present                         | 84 (29.5)       | 54 (27.7)              | 30 (33.3)               |         |
| **Satellite nodules**           |                 |                        |                         | 0.782   |
| Absent                          | 270 (94.7)      | 184 (94.4)             | 86 (95.6)               |         |
| Present                         | 15 (5.3)        | 11 (5.6)               | 4 (4.4)                 |         |
| **Peritumoral hypointensity**   |                 |                        |                         | 1.000   |
| Absent                          | 257 (90.2)      | 176 (90.3)             | 81 (90)                 |         |
| Present                         | 28 (9.8)        | 19 (9.7)               | 9 (10)                  |         |
| **MVI**                         |                 |                        |                         | 0.668   |
| Absent                          | 209 (73.3)      | 142 (72.8)             | 67 (74.4)               |         |
| Present                         | 75 (26.3)       | 53 (27.2)              | 22 (24.4)               |         |
| **Tumor number**                |                 |                        |                         | 0.478   |
### Characteristics

| Characteristics | Total (n = 285) | Training set (n = 195) | Validation set (n = 90) | P-value |
|----------------|----------------|------------------------|-------------------------|---------|
| 1              | 223 (78.2)     | 155 (79.5)             | 68 (75.6)               |         |
| 2              | 47 (16.5)      | 30 (15.4)              | 17 (18.9)               |         |
| 3              | 13 (4.6)       | 8 (4.1)                | 5 (5.6)                 |         |
| 4              | 2 (0.7)        | 2 (1.0)                | 0                       |         |

**Note:** Data are presented as number (%) or median (interquartile range, IQR); NE, neutrophil count; HBV, hepatitis B virus; AFP, alpha-fetoprotein; ALT, alanine amino-transferase; AST, aspartate amino-transferase; GGT, glutamyltranspeptadase; BCLC, barcelona clinic liver cancer; AP, arterial phase; Type 1, a homogeneous enhancement pattern with no increased arterial blood flow; Type 2, a homogeneous enhancement with increased arterial blood flow; Type 3, a heterogeneous enhancement included non-enhanced areas; Type 4, a heterogeneous enhancement pattern with irregular ring-like structures; Type 5, a heterogeneous and hypointense; MVI, microvascular invasion.

### DL signature development and validation

In total, 99 DL features were selected from AP images using RFE, 93 DL features were screened from PVP images using Relief, and 99 DL features were identified from HBP images using Relief. All the optimal single-layered DL signatures were constructed using GP classifiers. All DL signatures showed significant differences between the two groups (all P values < 0.05) in the training set (*Supplementary Table 3*). Collinearity analysis was performed for variables with P-value < 0.10 after univariate analysis. The tolerance range of each variable was 0.748-0.963, and the VIF range was 1.038-1.337, indicating no collinearity between the variables (*Supplementary Table 4*).

The AP-based DL signature achieved an AUC of 0.882 (95% confidence interval [CI]: 0.823-0.941) in the training set and 0.826 (95% CI: 0.755-0.897) in the validation set (*Table 2*); the PVP-based DL signature yielded an AUC of 0.822 (95% CI: 0.753-0.891) in the training set and 0.854 (95% CI: 0.780-0.928) in the validation set (*Table 2*); while the HBP-based DL signature obtained an AUC of 0.909 (95% CI: 0.860-0.958) in the training set and 0.888 (95% CI: 0.833-0.943) in the validation set (*Table 2*). The performance of the mp-MR-based DL signature improved in both training and validation sets, with an AUC of 0.929 (95% CI: 0.829-0.966) in the training set and 0.894 (95% CI: 0.811-0.977) in the validation set (*Table 2*).
Table 2
Predictive performance of different DL signature in training and validation sets

| Datasets    | Phase         | Sensitivity (%) | Specificity (%) | Accuracy (%) | AUC (95%CI)       |
|-------------|---------------|-----------------|-----------------|--------------|------------------|
| Training set| AP DL signature | 76.4            | 84.3            | 82.1         | 0.882 (0.823-0.941) |
|             | PVP DL signature | 67.3            | 85.0            | 80.0         | 0.822 (0.753-0.891) |
|             | HBP DL signature | 78.2            | 90.7            | 87.2         | 0.909 (0.860-0.958) |
|             | mp-MR DL signature | 76.4            | 93.6            | 88.7         | 0.929 (0.829-0.966) |
| Validation set | AP DL signature | 73.9            | 79.1            | 77.8         | 0.826 (0.755-0.897) |
|             | PVP DL signature | 73.9            | 86.6            | 83.3         | 0.854 (0.780-0.928) |
|             | HBP DL signature | 78.3            | 77.6            | 77.8         | 0.888 (0.833-0.943) |
|             | mp-MR DL signature | 78.3            | 89.6            | 86.7         | 0.894 (0.811-0.977) |

Note: AUC, area under the curve; CI, confidence interval; AP, arterial phase; DL, deep learning; PVP, portal venous phase; HBP, hepatobiliary phase; mp-MR, multiple sequences magnetic resonance.

Table 3
Multivariate regression analysis of the groups with and without early recurrence in the training set

| Variables         | Clinical nomogram | DL nomogram |DL nomogram |
|-------------------|-------------------|-------------|-------------|
|                   | Odd ratios (95%CI) | P-value     | Odd ratios (95%CI) | P-value |
| NE                | 0.70 (0.52-0.93)  | 0.014       | NA          | NA       |
| AST               | 3.07 (1.52-6.22)  | 0.002       | NA          | NA       |
| MVI               | 3.82 (1.87-7.80)  | <0.0001     | 3.06 (1.06-8.82) | 0.039   |
| Tumor number      | NA                | NA          | 3.61 (1.64-7.92) | 0.001   |
| mp-MR DL signature | NA                | NA          | 1.61×10^{15} (5.63×10^{10}-4.63×10^{19}) | <0.0001 |

Note: DL, deep learning; NE, neutrophil count; NA, not applicable; AST, aspartate amino-transferase; MVI, microvascular invasion; mp-MR, multiple sequences magnetic resonance.
Establishment of clinical nomogram and DL nomogram

After univariate analysis, seven clinical factors including neutrophil count, AFP, AST, GGT, BCLC stage, MVI, and tumor number (P<0.10) were candidates for multivariate analysis (Supplementary Table 2); however, only neutrophil count, AST, and MVI were identified as independent risk factors for early recurrence of HCC. The clinical nomogram was built based on the three indicators (Fig. 2a), achieving an AUC of 0.751 (95% CI: 0.674-0.827) in the training set and 0.712 (95% CI: 0.582-0.841) in the validation set (Table 4, Fig. 2c-d).

| Model                   | Training set | Validation set |
|-------------------------|--------------|----------------|
|                         | AUC (95%CI)  | Sensitivity (%) | Specificity (%) | AUC (95%CI)  | Sensitivity (%) | Specificity (%) |
| Clinical nomogram       | 0.751 (0.674-0.827) | 63.0          | 80.9          | 0.712 (0.582-0.841) | 56.5          | 84.8          |
| mp-MR DL signature      | 0.929 (0.829-0.966) | 76.4          | 93.6          | 0.894 (0.811-0.977) | 78.3          | 89.6          |
| DL nomogram             | 0.949 (0.919-0.980) | 90.7          | 85.5          | 0.908 (0.841-0.976) | 82.6          | 84.8          |

**Note:** AUC, area under the curve; CI, confidence interval; mp-MR, multiple sequences magnetic resonance; DL, deep learning.

After that, the above-mentioned seven clinical factors together with mp-MR-based DL signature were included in the multivariate analysis, and tumor number, MVI, and mp-DL signature were identified as independent risk factors for early recurrence (P < 0.05). We developed a DL nomogram incorporating the tumor number, MVI, and mp-MR-based DL signature (Fig. 2b), which significantly outperformed the clinical nomogram, yielding an AUC of 0.949 (95% CI: 0.919-0.980) in the training set and 0.908 (95% CI: 0.841-0.976) in the validation set (Table 4, Fig. 2c-d). The Delong test showed a significant difference between the clinical nomogram and DL nomogram in the training set (P < 0.0001) and validation set (P = 0.002). Fig. 2e-f demonstrate good calibration of the DL nomogram. The DCA curve showed that the DL nomogram had a higher net benefit than the clinical nomogram (Fig. 2g-h).

**Discussion**
In this present study, we used the DL approach to explore the informative features from Gd-EOB-DTPA-enhanced MRI images that were associated with early recurrence of HCC and established three single-layered DL signatures and mp-MR DL signature fused with three-phase MR sequences. The results showed that the mp-MR DL signature was better than the three single-layered DL signatures. Subsequently, the DL nomogram was constructed by integrating tumor number, MVI, and the mp-MR DL signature, which achieved higher predictive accuracy and better net benefit than the clinical nomogram. This study demonstrated the incremental value of the DL signature as compared to the conventional clinical nomogram.

Some past studies have explored the feasibility of predicting early recurrence of HCC after surgery. Zhang et al. [19] developed nomograms containing clinicoradiological variables to predict the early recurrence in patients with HCC, achieving AUCs of 0.836-0.846 in the validation cohorts. Chan et al. [20] constructed a pre-operative model (ERASL-pre) and a post-operative model (ERASL-post) based on gender, AFP, MVI, tumor size, and tumor number, with AUCs ranging from 0.614 to 0.736 for the ERASL-pre and from 0.653 to 0.763 for the ERASL-post in the derivation cohorts from different countries or districts. In our study, we also built a clinical nomogram and the diagnostic efficacy was similar to previous studies, with an AUC of 0.751 and 0.712 in the training and validation sets, respectively. Although these data are commonly and easily available in clinical practice, the two major problems were non-uniformity of standardization of clinical indicators and the failure of traditional statistical methods to handle high-dimension features [21]. In the field of oncology, radiomics is a recently emerged technology that extracts a large number of quantitative image features from standard-of-care medical imaging using data-characterisation algorithms. Zhao et al. [22] constructed a nomogram by integrating radiomic score and clinicoradiological factors, with an AUC of 0.873. Kim et al. [23] developed a combined clinicopathologic-radiomic model via RF algorithm, which acquired a C-index of 0.716. Although these radiomic studies yielded good performance, they were limited to a small sample size and a single feature selection method. In the field of feature engineering, different machine learning based dimensionality reduction techniques have distinct mathematical senses and inherent limitations; thus, multiple algorithms should be combined to select robust features [24]. Prior studies also proved that different dimensional reduction methods combined with several machine learning methods could maximize the diagnostic performance of the model. Dai et al. [25] found that feature selection and modeling methods could have potential effects on prediction models. The optimal radiomic model for MVI evaluation was constructed by gradient boosting decision tree (GBDT) classifier, which outperformed logistic regression, SVM, and RF. Ni et al. [26] identified LASSO plus GBDT as the optimal combination for predicting MVI in HCC patients from 21 combination methods (three feature selection methods and seven classification methods). In our study, we compared three feature selection methods together with five classification methods to determine the best combination and found that RFE or Relief combined with GP classifier obtained the optimal performance in building DL signatures.

Recently, artificial intelligence has emerged as an effective tool to demonstrate multi-modal patient data [27]. Given the fact that radiomics comes with time-consuming progress of features engineering, we used the VGGnet-19, which can generally perform a more robust image task through automatic analysis
without export's intervention [28]. Gd-EOB-DTPA-enhanced MRI could better capture the perfusion and functional alterations, and be more sensitive and accurate in the detection of HCC. Hence, we chose Gd-EOB-DTPA-enhanced MRI instead of CT images to predict the early recurrence in patients with HCC. Liu et al. [29] also revealed that Gd-EOB-DTPA-enhanced MRI had a significantly higher sensitivity and overall accuracy for HCC especially small lesions than multiphasic CT without substantial loss of specificity. Wei et al. [30] explored the validity of DL in predicting MVI of HCC using CT and Gd-EOB-DTPA-enhanced MRI to train two DL models and suggested that the EOB-MRI based model achieved a better prediction result (AUC = 0.812) than the CT-based model (AUC = 0.736). In addition, the results of our study demonstrated that the HBP signature yielded higher AUC and sensitivity than AP signature and PVP signature, whereas lower specificity and accuracy than PVP signature. Well-defined tumor margins on HBP images allow for a more accurate tumor delineation than that on AP and PVP images, in which the tumor margins could be affected by peritumoral enhancement, capsule appearance, and hypodense halo. Furthermore, this current study also showed the complementary role of MR sequences in DL analysis.

Our study also has several limitations. First, the retrospective nature of this study may induce inevitable selection bias. Second, our study was conducted in two centers, more institutions for external validation are required in the future. Finally, the value of the DL model for improving long-term survival in HCC patients remains unclear, the differences between DL model-assisted and non-assisted practices warrant further study to prove the clinical application of the DL model.

In conclusion, a DL-nomogram based on clinical factors and Gd-EOB-DTPA-enhanced MRI biomarkers could be applied to accurate prediction of early recurrence after surgery and postoperative surveillance for HCC patients following surgical resection.

**Abbreviations**

HCC: Hepatocellular carcinoma

MVI: microvascular invasion

Gd-EOB-DTPA: Gadolinium-ethoxybenzyl-diethylenetriamine pentaacetic acid

MRI: Magnetic resonance imaging

HBP: Hepatobiliary phase

DL: Deep learning

CNN: Convolutional neural network

AFP: α-fetoprotein

ALT: Alanine aminotransferase
AST: aspartate aminotransferase
GGT: γ-Glutamyltranspeptidase
BCLC: Barcelona Clinic Liver Cancer
AP: Arterial phase
PVP: Portal venous phase
RFE: recursive feature elimination
RF: random forest
SVM: support vector machine
LASSO: least absolute shrinkage and selection operator logistic regression
GP: Gaussian process
VIF: Variance inflation factor
ROC: Receiver operating characteristic curve
AUC: Area under the curve
DCA: Decision curve analysis
mp-MR: Multiple sequences magnetic resonance

Declarations

Funding

This work was supported by the Science and Technology Planning Project of Guangdong Province of China [2017A020215010].

Author Contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Meng Yan, Xiao Zhang, Zhijun Geng, Xinming Li, Zhendong Qi, Ting Lin, and Qiying Ke. Administrative support provided by Xianyue Quan, Chuanmiao Xie, Wei Yang, Shutong Wang and Shuixing Zhang. The first draft of the manuscript was written by Meng yan, Bin Zhang, Xiao Zhang and Xinming Li. All authors read and approved the final manuscript.

Compliance with Ethical Requirements
Conflict of Interest statements

Meng Yan, Xiao Zhang, Bin Zhang, Zhijun Geng, Chuanmiao Xie, Wei Yang, Shuixing Zhang, Zhendong Qi, Ting Lin, Qiying Ke, Xinming Li, Shutong Wang and Xianyue Quan declares that they have no conflict of interesting.

Ethical approval

The study was conducted in accordance with the ethics principles of the Declaration of Helsinki in 1975, as revised in 2008. Ethic approval for this study was obtained from the Ethics Committees of Zhujiang Hospital of Southern Medical University (2020-KY-094-01) and Sun Yat-Sen University Cancer Center (SL-B2021-214-02), and the requirement for informed consent was waived.

References

1. Marasco G, Colecchia A, Colli A, Ravaíoli F, Casazza G, Bacchi Reggiani ML et al. Role of liver and spleen stiffness in predicting the recurrence of hepatocellular carcinoma after resection. J Hepatol. 2019;70(3):440-8.
2. Ishizawa T, Hasegawa K, Aoki T, Takahashi M, Inoue Y, Sano K et al. Neither multiple tumors nor portal hypertension are surgical contraindications for hepatocellular carcinoma. Gastroenterology. 2008;134(7):1908-16.
3. Tung-Ping Poon R, Fan ST, Wong J. Risk factors, prevention, and management of postoperative recurrence after resection of hepatocellular carcinoma. Annals of surgery. 2000;232(1):10-24.
4. Portolani N, Coniglio A, Ghidoni S, Giovanelli M, Benetti A, Tiberio GA et al. Early and late recurrence after liver resection for hepatocellular carcinoma: prognostic and therapeutic implications. Annals of surgery. 2006;243(2):229-35.
5. Lee S, Kang TW, Song KD, Lee MW, Rhim H, Lim HK et al. Effect of Microvascular Invasion Risk on Early Recurrence of Hepatocellular Carcinoma After Surgery and Radiofrequency Ablation. Annals of surgery. 2021;273(3):564-71.
6. Hirokawa F, Hayashi M, Asakuma M, Shimizu T, Inoue Y, Uchiyama K. Risk factors and patterns of early recurrence after curative hepatectomy for hepatocellular carcinoma. Surg Oncol. 2016;25(1):24-9.
7. Lee YJ, Lee JM, Lee JS, Lee HY, Park BH, Kim YH et al. Hepatocellular carcinoma: diagnostic performance of multidetector CT and MR imaging—a systematic review and meta-analysis. Radiology. 2015;275(1):97-109.
8. Lee S, Kim SH, Lee JE, Sinn DH, Park CK. Preoperative gadoxetic acid-enhanced MRI for predicting microvascular invasion in patients with single hepatocellular carcinoma. J Hepatol. 2017; 67(3):526-34.
9. Zhang W, Fang M, Dong D, Wang X, Ke X, Zhang L et al. Development and validation of a CT-based radiomic nomogram for preoperative prediction of early recurrence in advanced gastric cancer.
Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology. 2020;145:13-20.

10. Liu D, Liu F, Xie X, Su L, Liu M, Xie X et al. Accurate prediction of responses to transarterial chemoembolization for patients with hepatocellular carcinoma by using artificial intelligence in contrast-enhanced ultrasound. European radiology. 2020;30(4):2365-76.

11. Murphy A, Skalski M, Gaillard F. The utilisation of convolutional neural networks in detecting pulmonary nodules: a review. The British journal of radiology. 2018;91(1090):20180028.

12. Roayaie S, Blume IN, Thung SN, Guido M, Fiel MI, Hiotis S et al. A system of classifying microvascular invasion to predict outcome after resection in patients with hepatocellular carcinoma. Gastroenterology. 2009;137(3):850-5.

13. Kawamura Y, Ikeda K, Seko Y, Hosaka T, Kobayashi M, Saitoh S et al. Heterogeneous type 4 enhancement of hepatocellular carcinoma on dynamic CT is associated with tumor recurrence after radiofrequency ablation. AJR Am J Roentgenol. 2011;197(4):W665-73.

14. Kawamura Y, Ikeda K, Hirakawa M, Yatsuji H, Sezaki H, Hosaka T et al. New classification of dynamic computed tomography images predictive of malignant characteristics of hepatocellular carcinoma. Hepatology research : the official journal of the Japan Society of Hepatology. 2010;40(10):1006-14.

15. Shimizu R, Tamai H, Mori Y, Shingaki N, Maeshima S, Nuta J et al. The arterial tumor enhancement pattern on contrast-enhanced computed tomography is associated with primary cancer death after radiofrequency ablation for small hepatocellular carcinoma. Hepatology international. 2016;10(2):328-39.

16. Huang M, Liao B, Xu P, Cai H, Huang K, Dong Z et al. Prediction of Microvascular Invasion in Hepatocellular Carcinoma: Preoperative Gd-EOB-DTPA-Dynamic Enhanced MRI and Histopathological Correlation. Contrast Media Mol Imaging. 2018;2018:9674565.

17. Ji G-W, Zhu F-P, Xu Q, Wang K, Wu M-Y, Tang W-W et al. Radiomic Features at Contrast-enhanced CT Predict Recurrence in Early Stage Hepatocellular Carcinoma: A Multi-Institutional Study. Radiology. 2020;294(3):568-79.

18. Kim KA, Kim MJ, Jeon HM, Kim KS, Choi JS, Ahn SH et al. Prediction of microvascular invasion of hepatocellular carcinoma: usefulness of peritumoral hypointensity seen on gadoxetate disodium-enhanced hepatobiliary phase images. Journal of magnetic resonance imaging : JMRI. 2012;35(3):629-34.

19. Zhang XP, Chen ZH, Zhou TF, Li LQ, Chen MS, Wen TF et al. A nomogram to predict early postoperative recurrence of hepatocellular carcinoma with portal vein tumour thrombus after R0 liver resection: A large-scale, multicenter study. European journal of surgical oncology: the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology. 2019;45(9):1644-51.

20. Chan AWH, Zhong J, Berhane S, Toyoda H, Cucchetti A, Shi K et al. Development of pre and post-operative models to predict early recurrence of hepatocellular carcinoma after surgical resection. J Hepatol. 2018;69(6):1284-93.
21. Zou ZM, Chang DH, Liu H, Xiao YD. Current updates in machine learning in the prediction of therapeutic outcome of hepatocellular carcinoma: what should we know? Insights into imaging. 2021;12(1):31.

22. Zhao Y, Wu J, Zhang Q, Hua Z, Qi W, Wang N et al. Radiomics Analysis Based on Multiparametric MRI for Predicting Early Recurrence in Hepatocellular Carcinoma After Partial Hepatectomy. Journal of magnetic resonance imaging : JMRI. 2021;53(4):1066-79.

23. Kim S, Shin J, Kim DY, Choi GH, Kim MJ, Choi JY. Radiomics on Gadoxetic Acid-Enhanced Magnetic Resonance Imaging for Prediction of Postoperative Early and Late Recurrence of Single Hepatocellular Carcinoma. Clinical cancer research : an official journal of the American Association for Cancer Research. 2019;25(13):3847-55.

24. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J et al. Radiomics: the bridge between medical imaging and personalized medicine. Nature reviews Clinical oncology. 2017;14(12):749-62.

25. Dai H, Lu M, Huang B, Tang M, Pang T, Liao B et al. Considerable effects of imaging sequences, feature extraction, feature selection, and classifiers on radiomics-based prediction of microvascular invasion in hepatocellular carcinoma using magnetic resonance imaging. Quantitative imaging in medicine and surgery. 2021;11(5):1836-53.

26. Ni M, Zhou X, Lv Q, Li Z, Gao Y, Tan Y et al. Radiomics models for diagnosing microvascular invasion in hepatocellular carcinoma: which model is the best model? Cancer imaging : the official publication of the International Cancer Imaging Society. 2019;19(1):60.

27. Ahn JC, Connell A, Simonetto DA, Hughes C, Shah VH. Application of Artificial Intelligence for the Diagnosis and Treatment of Liver Diseases. Hepatology. 2021;73(6):2546-63.

28. Sung YS, Park B, Park HJ, Lee SS. Radiomics and deep learning in liver diseases. Journal of gastroenterology and hepatology. 2021;36(3):561-8.

29. Liu X, Jiang H, Chen J, Zhou Y, Huang Z, Song B. Gadoxetic acid disodium-enhanced magnetic resonance imaging outperformed multidetector computed tomography in diagnosing small hepatocellular carcinoma: A meta-analysis. Liver transplantation : official publication of the American Association for the Study of Liver Diseases and the International Liver Transplantation Society. 2017;23(12):1505-18.

30. Wei J, Jiang H, Zeng M, Wang M, Niu M, Gu D et al. Prediction of Microvascular Invasion in Hepatocellular Carcinoma via Deep Learning: A Multi-Center and Prospective Validation Study. Cancers. 2021;13(10).

Figures
Figure 1

The workflow of deep learning analysis. MR, magnetic resonance; ROI, region of interest; RFE, recursive feature elimination; ROC, receiver operating characteristic curve; DL, deep learning.
Figure 2

a Clinical nomogram was developed with NE, AST, and MVI. b DL nomogram was developed with MVI, tumor number, and mp-MR DL signature. c The ROC curves for clinical nomogram and DL nomogram in the training set. d The ROC curves for clinical nomogram and DL nomogram in the validation set. e Calibration curves for the clinical nomogram and DL nomogram in the training set. f Calibration curves for the clinical nomogram and DL nomogram in the validation set. The y-axis represents the actual early
recurrence rate, the x-axis represents the predicted early recurrence and the diagonal dashed line represents the ideal prediction by a perfect model. **g** Decision curves analysis for the clinical nomogram and DL nomogram in the training set. **h** Decision curves analysis for the clinical nomogram and DL nomogram in the validation set. The y-axis represents the net benefits, and the x-axis represented the threshold probability. The DL nomogram (red line) had a good net benefit compared with the clinical nomogram (blue line), simple strategies such as follow-up of all patients (grey line) or no patients (horizontal black line) across the majority range of threshold probabilities. NE, neutrophil count. AST, aspartate amino-transferase. MVI, microvascular invasion; DL, deep learning; mp-MR, multiple sequences magnetic resonance; ROC, receiver operator characteristic

### Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementaryFig.12.pdf](#)
- [supplementarymaterials.pdf](#)