Hepatology International (2020) 14:567–576
https://doi.org/10.1007/s12072-020-10046-w

ORIGINAL ARTICLE

Machine learning-based development and validation of a scoring system for progression-free survival in liver cancer

Xiaoli Liu[1] · Yixin Hou[1] · Xinhui Wang[1] · Lihua Yu[1] · Xianbo Wang[1] · Li Jiang[2] · Zhiyun Yang[1]

Received: 30 December 2019 / Accepted: 10 April 2020 / Published online: 18 June 2020
© Asian Pacific Association for the Study of the Liver 2020

Abstract

Object Disease progression is an important factor affecting the long-term survival in hepatocellular carcinoma (HCC). The progression-free survival (PFS) has been used as a surrogate endpoint for overall survival (OS) in many solid tumors. However, there were few models to predict the PFS in HCC patients. This study aimed to explore the prognostic factors that affect the PFS in HCC and establish an individualized prediction model.

Methods We included 2890 patients with hepatitis B-related HCC hospitalized at Beijing Ditan Hospital, Capital Medical University and randomly divided into training and validation cohort. Cox multivariate regression was used to analyze independent risk factors affecting the 1-year PFS of HCC, and an artificial neural networks (ANNs) model was constructed. C-index, calibration curve, and decision curve analysis were used to evaluate the performance of the model.

Results The median survival time was 26.2 m (95% CI: 24.08–28.32) and the 1-year PFS rate was 52.3% in whole study population. Cox multivariate regression showed smoking history, tumor number ≥ 2, tumor size ≥ 5 cm, portal vein tumor thrombus, WBC, NLR, γ-GGT, ALP, and AFP ≥ 400 ng/mL were risk factors for 1-year progression-free survival, while albumin and CD4 T cell counts were protective factors in HCC patients. A prediction model for 1-year PFS was constructed (https://lixuan.me/annmodel/myg-v3/). The ANNs model's ability to predict 1-year PFS had an area under the receiver operating characteristic curve (AUROC) of 0.866 (95% CI 0.848–0.884) in HCC patients, which was higher than predicted by TNM, BCLC, Okuda, CLIP, CUPI, JIS, and ALBI scores ($p < 0.0001$). In addition, the ANNs model could also estimate the probability of 1-year OS and presented a higher AUROC value, 0.877 (95% CI 0.858–0.895), than those other models. All patients were divided into high-, medium-, and low-risk groups, according to the ANNs model scores. Compared with the hazard ratios (HRs) of PFS and OS in low-risk group, those in the high-risk group were 26.42 (95% CI 18.74–37.25; $p < 0.0001$) and 11.26 (95% CI 9.11–13.93; $p < 0.0001$), respectively.

Conclusion The ANNs model has good individualized prediction performance and may be helpful to evaluate the probability of progression-free survival in HCC during clinical practice.

Keywords Machine learning · Artificial neural networks · HCC · Progression-free survival · Prognosis · CD4 · Risk factor · Hepatitis B virus · Immunology · Circulating T cells

Introduction

Liver cancer is the fourth leading cause of death among all cancers worldwide [1]. Although the survival time of hepatocellular carcinoma (HCC) has been effectively prolonged owing to the discovery of early monitoring indicators, regular examination and advances in medical technology, the death rate of HCC is still increasing due to population growth and aging. The years of life lost (YLLs) caused by HCC will increase 69.6% by 2040 compared with the current level [2].

Xiaoli Liu and Yixin Hou have contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s12072-020-10046-w) contains supplementary material, which is available to authorized users.

✉ Li Jiang
movdush@139.com

✉ Zhiyun Yang
yangzhiyun2016@163.com

Extended author information available on the last page of the article

There are various treatment methods for HCC at different stages, but the recurrence and metastasis after treatment are still serious clinical problems. Some studies have reported that patients with HCC still have a recurrence rate of 8–20% after liver transplantation, and the median survival time is only 8.7 months once patients experienced recurrence [3]. The early (< 1 year) and late (≥ 1 year) intrahepatic recurrence rates are 19–38% and 41–70% in patients after liver resection [4, 5]. Radiofrequency therapy is an alternative therapy for unresectable HCC, and the 3-year distant recurrence rate in patients after treatment is 63.3% [6]. Therefore, the disease progression is still an important cause of death in HCC patients.

At present, models used to predict the recurrence of HCC are mostly focused on the resection or liver transplantation [7, 8], and there were few models for progression of the middle and advanced HCC with other treatments. Progression-free survival (PFS) more often used to replace the overall survival in many solid tumors [9]. There were also researches that suggest PFS as an appropriate surrogate of survival in advanced HCC [10]. PFS is more attractive than OS because it measured earlier at equivalent durations of follow-up, and is not affected by competing risks of death [11]. However, there were few studies that predict the PFS of HCC patients.

As a type of machine learning, the artificial neural networks (ANNs) is essentially a mathematical model driven by the biological nervous system (similar to information processing by brain neurons) [12]. ANNs have been widely used in medical decision making, and are advantageous because they can complete statistical analysis of linear, logistic, or nonlinear complex relationships [13]. The purpose of this study is to use ANNs to construct an early-stage warning model for predicting the 1-year PFS in patients with HBV–HCC to accurately identify high-risk groups that may experience disease progression and initiate intervention at the early stage to reduce the mortality rate in patients.

Materials and methods

Patients

A total of 2890 patients with first-diagnosed HBV-related primary liver cancer who were hospitalized in Beijing Ditan Hospital, Capital Medical University from January 2008 to December 2016 were retrospectively included. This study was approved by the Ethics Committee of Beijing Ditan Hospital. Inclusion criteria: (1) patients diagnosed with primary liver cancer; (2) aged 18–75 years; (3) HBsAg positive lasting > 6 months. Exclusion criteria were shown in Fig S1. We randomized the 2117 patients that finally enrolled, with 70% in the training cohort ($n = 1480$) and 30% in the validation cohort ($n = 637$) (Fig. S1). The diagnostic criteria for HCC was consistent with our previous study [14].

Follow-up and endpoint

All patients underwent CT, MRI, ultrasonography, or serum AFP tests every 3 months. When the serum AFP level of the patient increased or new intrahepatic nodules were found by ultrasonography, dynamic CT or MRI was used to determine whether the HCC had progressed. The definition of progression conformed with the mRECIST criteria [15]. The occurrence of vascular metastasis or extrahepatic diffusion was also considered as progression. Progression-free survival was defined as the time from baseline to progression or death, whichever came first. Survival time was defined as the time from enrolment of the patient to death or to December 31, 2018, whichever was earlier.

Statistical analysis

SPSS 21.0 was used for data analysis. Quantitative data were compared using the Student's *t* test or Mann–Whitney *U* test. Qualitative data were compared by Fisher's exact test or χ^2 test. Cox univariate and multivariate analysis (forward selection, maximum likelihood ratio method) were used to screen out the risk factors affecting the 1-year PFS of HCC. The ANNs model was constructed with the factors related with the PFS of HCC using Mathematica 11.1.1 for Microsoft windows (64-bit). The final established model was compared with the following prognosis models: TNM [16], BCLC [17], Okuda [18], CLIP [19], CUPI [20], JIS [21], and ALBI score [22]. The discrimination degree of the models was tested using the concordance index (C-index). To test the calibration degree of the model, the Hosmer–Lemeshow test was applied and calibration curves were plotted. Decision-curve analysis (DCA) was used to compare the clinical net benefits and performance improvement between this model and the above models. Data analysis was conducted using R language version 3.3.2, with rms, survival, and rmda packages. For all tests, $p < 0.05$ was considered as indicating a statistically difference.

Results

Baseline characteristics

We enrolled 2117 patients between 2008 and 2016 and randomly divided them into the training ($n = 1480$) and validation ($n = 637$) datasets. In the training group, 434 (29.3%) patients developed disease progression within 1 year, 174 (40.1%) of which died within 1 year and 310 (71.4%) died within 3 years after progression. Among 784 1-year

progression-free survival patients, only 173 (22.1%) died within 3 years. In the validation group, 181 (28.4%) patients developed disease progression within 1 year, 93 (51.4%) of which died within 1 year and 134 (74.0%) died within 3 years after progression. Among 327 1-year progression-free survival patients, only 82 (25.1%) died within 3 years (Fig. S1). The baseline characteristics of the two groups were comparable (Table 1). Among all patients with HBV-related HCC, 1848 (87.3%) received antiviral therapy, 1388 (65.6%) achieved serological conversion, and 1148 (54.2%) achieved HBV-DNA < 500 IU/mL. 68.3% HCC patients received local treatments such as TACE or RFA, while the remaining 9.3% received radical resection and 22.4% received palliative treatment.

Survival analysis

For the whole study population, the median survival time was 26.2 m (95% CI: 24.08–28.32); the 1-year PFS rate was 52.3%; the 1-year, 3-year, and 5-year OS rates were 66.4%, 45.1%, and 26.6%, respectively (Fig. 2a, b). In the training group, the median survival time was 27.20 m (95% CI: 24.48–29.92); the 1-year PFS rate was 52.8%; the 1-year, 3-year, and 5-year OS rates were 66.3%, 46.3%, and 26.5%, respectively. In the validation group, the median survival time was 24.53 m (95% CI: 21.52–27.55); the 1-year PFS rate was 51.3%; the 1-year, 3-year, and 5-year OS rates were 66.7%, 47.3%, and 26.8%, respectively. There was no statistically significant difference between the two groups.

Development of the prognosis model based on the artificial neural network

As shown in Table S1, we found through the cox univariate and multivariate regression analysis that smoking history (adjusted HR = 1.188, 95% CI 1.017–1.386, $p = 0.029$), tumor size ≥ 5 cm (adjusted HR = 2.357, 95% CI 1.993–2.788, $p < 0.0001$), tumor number ≥ 2 (adjusted HR = 1.243, 95% CI 1.066–1.449, $p = 0.001$), portal vein tumor thrombus (adjusted HR = 1.776, 95% CI 1.5–2.104, $p < 0.001$), WBC (adjusted HR = 1.053, 95% CI 1.02–1.087, $p = 0.001$), NLR (adjusted HR = 1.031, 95% CI 1.012–1.051, $p = 0.002$), γ-GGT (adjusted HR = 1.002, 95% CI 1.001–1.002, $p < 0.0001$), ALP (adjusted HR = 1.001, 95% CI 1.001–1.002, $p < 0.0001$), and AFP ≥ 400 ng/mL (adjusted HR = 1.723, 95% CI 1.456–2.026, $p < 0.0001$) were risk factors for 1-year PFS in patients with HCC, while ALB (adjusted HR = 0.971, 95% CI 0.959–0.984, $p < 0.0001$) and CD4 T cell counts (adjusted HR = 0.998, 95% CI 0.998–0.998, $p < 0.0001$) were protective factors. These indicators were also included and used to construct the ANNs model (https://lixuan.me/annmodel/myg-v3/).

Multilayer perceptron (MLP) is the most common ANN structure. Its basic components include the input layer, hidden layer, and output layer [12]. The input layer includes clinical or biochemical parameters, and the output layer includes the corresponding prognosis outcomes. As shown in Fig. 1a, neurons are interconnected by weighted links, with a total of 11 input neurons and two output neurons. To improve the performance of MLP, we added three hidden layers after many rounds of debugging and testing.

Discrimination and calibration of the ANN model

For the 1-year PFS and 1-year OS in patients with HCC in the training cohort, the areas under the ROC curve (AUROC) were 0.866 (95% CI 0.848–0.884) and 0.877 (95% CI 0.858–0.895), respectively; the C-index values were 0.782 (95% CI 0.767–0.797) and 0.824 (95% CI 0.808–0.840), respectively (Table 2). The AUROC values and C-index values of the ANNs model were all higher than those of BCLC, TNM, Okuda, CUPI, CLIP, JIS, ALBI, MELD, and MELD-Na models ($p < 0.0001$). For the 1-year PFS and 1-year OS in patients with HCC in the validation cohort, the AUROC were 0.730 (95% CI 0.690–0.770) and 0.804 (95% CI 0.765–0.843), respectively (Table 2). We further compared AUC and C-index values among different sex, age group, AFP levels, Child–Pugh grades, and different treatment methods and found that the ANNs model was superior to other models (Table S2, 3).

The predicted 1-year PFS probability, 1-year OS probability, and the corresponding actual observation probabilities were shown in the calibration curves, respectively (Fig. 1). The ANNs model had a good fit in 1-year PFS (Fig. 1b) and OS (Fig. 1c) in both the training cohort and validation cohort. In decision curve analysis, compared with BCLC, TNM, Okuda, CUPI, CLIP, and JIS models, our ANNs model demonstrated a significant net benefit, with improved performance in prognosis evaluation for 1-year PFS (Fig. 1d) and OS (Fig. 1e) in both the training cohort and validation cohort. This indicates that the ANNs model has better clinical practicability than other models.

Application of the ANNs model for risk stratification

We divided all patients into three strata according to the upper quartiles and lower quartiles of the ANNs model scores: Strata 1, low risk; Strata 2, medium risk; Strata 3, high risk. In the training cohort, taking Strata 1 as the reference, the hazard ratios (HRs) of PFS of Strata 2 and Strata 3 were 5.25 (95% CI 3.73–7.38) and 26.42 (95% CI 18.74–37.25) ($p < 0.0001$), respectively (Fig. 2c); the HRs of OS of Strata 2 and Strata 3 were 2.60 (95% CI 2.12–3.17) and 11.26 (95% CI 9.11–13.93) ($p < 0.0001$), respectively (Fig. 2e). In the validation cohort, there were

Table 1 Demographic data and clinical characteristics of patients with hepatocellular carcinoma

	Total ($n=2117$)	Training cohort $n=1480$ (%)	Validation cohort $n=637$ (%)	p values
Patients background				
Age, years (mean ± SD)	55.65 ± 10.2	55.47 ± 10.1	56.08 ± 10.43	0.338
Gender (male/female)	1668/449 (78.8%/21.2%)	1174/306 (79.3%/20.7%)	494/143 (77.6%/22.4%)	0.36
Family history of HCC (Yes/No)	80/2037(3.8%/96.2%)	50/1430 (3.4%/96.6%)	30/607 (4.7%/95.3%)	0.141
History of smoking (Yes/No)	841/1276 (39.7%/60.3%)	589/891 (39.8%/60.2%)	252/385 (39.6%/60.4%)	0.919
History of alcohol use (Yes/No)	773/1344 (36.5%/63.5%)	528/952 (35.7%/64.3%)	245/392 (38.5%/61.5%)	0.222
Diabetes (Yes/No)	417/1700 (19.7%/80.3%)	286/1194 (19.3%/80.7%)	131/506 (20.6%/79.4%)	0.51
Hypertension (Yes/No)	504/1613 (23.8%/76.2%)	365/1115 (24.7%/75.3%)	139/498(21.8%/78.2%)	0.159
Hyperlipidemia (Yes/No)	154/1963 (7.3%/92.7%)	112/1368 (7.6%/92.4%)	42/595 (6.6%/93.4%)	0.429
Coronary artery disease (Yes/No)	45/2072 (2.1%/97.9%)	37/1443 (2.5%/97.5%)	8/629 (1.3%/98.7%)	0.069
Cirrhosis (Yes/No)	1956/161 (92.4%/7.6%)	1362/118 (92.0%/8.0%)	594/43 (93.2%/6.8%)	0.33
HBV-related indicators at baseline				
HBeAg (Positive/Negative)	729/1388 (34.4%/65.6%)	509/971 (34.4%/65.6%)	220/417 (34.5%/65.5%)	0.949
HBV-DNA, IU/ml (≥ 500/ < 500)	969/1148 (45.8%/54.2%)	675/805 (45.6%/54.4%)	294/343 (46.2%/53.8%)	0.817
Antiviral therapy (Yes/No)	1848/269 (87.3%/12.7%)	1296/184 (87.6%/12.4%)	552/85 (86.7%/13.3%)	0.564
Laboratory data				
White blood cells (10^9/L)	4.46 (3.20, 5.98)	4.49 (3.24, 6.01)	4.41 (3.03, 5.90)	0.226
NLR	2.43 (1.63, 3.93)	2.44 (1.64, 3.91)	2.42(1.60, 4.08)	0.999
Platelets (10^9/L)	93.1 (59.6, 143.55)	93.0 (61.0, 144.0)	94.0 (58.0, 141.25)	0.342
ALT (U/L)	33.4 (22.7, 54.3)	33.6 (22.8, 54.3)	32.8 (22.5, 54.15)	0.777
AST (U/L)	41.1 (27.7, 69.95)	41.05 (27.9, 69.2)	41.1 (27.5, 72.75)	0.701
Total bilirubin (μmol/L)	18.5 (12.6, 30.75)	18.3 (12.6, 30.3)	19.1 (12.8, 32.15)	0.414
γ-GGT (U/L)	60.6 (30.25, 126.3)	62.45 (30.6, 131.0)	56.5 (29.0, 111.6)	0.117
Albumin (g/L)	34.30 ± 6.31	33.62 ± 6.18	35.38 ± 6.37	0.369
ALP (U/L)	96.7 (71.5, 133.7)	97.5 (72.1, 133.8)	95.9 (70.6, 131.6)	0.422
Cholinesterase (U/L)	4446 (2840, 6358.5)	4524.5 (2898.5, 6417.75)	4242.0 (2714.0, 6271.0)	0.065
Creatinine (μmol/L)	66.0 (57.0, 76.5)	66.0 (57.0, 76.0)	66.4 (57.45, 77.0)	0.385
MELD scores	5.24 (2.27, 8.34)	5.12 (2.12, 8.23)	5.42 (2.72, 8.46)	0.122
Prothrombin activity (%)	76.12 ± 18.14	76.66 ± 18.30	74.86 ± 17.71	0.035
INR	1.17 ± 0.23	1.16 ± 0.24	1.18 ± 0.22	0.166
AFP (ng/ml) (< 400/ ≥ 400)	1555/562 (73.5%/26.5%)	1069/411 (72.2%/27.8%)	486/151 (76.3%/23.6%)	0.052
CRP (mg/L)	16.4 (3.0, 19.7)	17.1 (3.0, 19.4)	13.0 (3.0, 20.0)	0.708
T cell counts (cells/μL)	740.5 (475.25, 1085.25)	743.0 (483.0, 1056.5)	735.0 (461.0, 1117.0)	0.876
CD4 T cell counts (cells/μL)	448.0 (280.0, 653.0)	445.0 (280.5, 649.5)	451.0 (279.75, 669.5)	0.842
CD8 T cell counts (cells/μL)	249.0 (151.0, 398.75)	252.0 (155.0, 397.0)	240.0 (147.0, 410.0)	0.794
Tumor-related indicators				
Tumor multiplicity (solitary/multiple)	1223/894 (57.8%/42.2%)	843/637 (57.0%/43.0%)	380/ 257 (59.7%/40.3%)	0.25
Tumor size, cm (< 5/ ≥ 5)	1480/637 (69.9%/30.1%)	1053/452 (70%/30%)	427/185 (69.8%/30.2%)	0.929
Portal vein tumor thrombus (Yes/No)	426/1691 (20.1%/79.9%)	301/1179 (20.3%/79.7%)	125/512 (19.6%/80.4%)	0.707
BCLC staging				
0–A	799/1318 (37.7%/62.3%)	546/934 (36.9%/63.1%)	253/384 (39.7%/60.3%)	0.219
B	666/1451 (31.5%/68.5%)	475/1005 (32.1%/67.9%)	191/446 (30.0%/70.0%)	0.338
C	391/1726 (18.5%/81.5%)	286/1194 (19.3%/80.7%)	105/532 (16.5%/83.5%)	0.122
D	261/1856 (12.3%/87.7%)	173/1307 (11.7%/88.3%)	88/549 (13.8%/86.2%)	0.172
Type of treatment				
Resection (Yes/No)	197/1920 (9.3%/90.7%)	128/1352 (8.6%/91.4%)	69/568 (10.8%/89.2%)	0.113
Minimally invasive (Yes/No)	1445/672 (68.3%/31.7%)	1029/451 (69.5%/30.5%)	416/221 (65.3%/34.7%)	0.056
Palliative (Yes/No)	475/1642 (22.4%/77.6%)	323/1157 (21.8%/78.2%)	152/485 (23.9%/76.1%)	0.303

SD standard deviation, *NLR* neutrophil–lymphocyte ratio, *ALT* alanine aminotransferase, *AST* aspartate aminotransferase, *γ-GGT* γ-glutamyl transferase, *ALP* alkaline phosphatase, *INR* international normalized ratio, *AFP* alpha-fetoprotein, *CRP* C-reactive protein

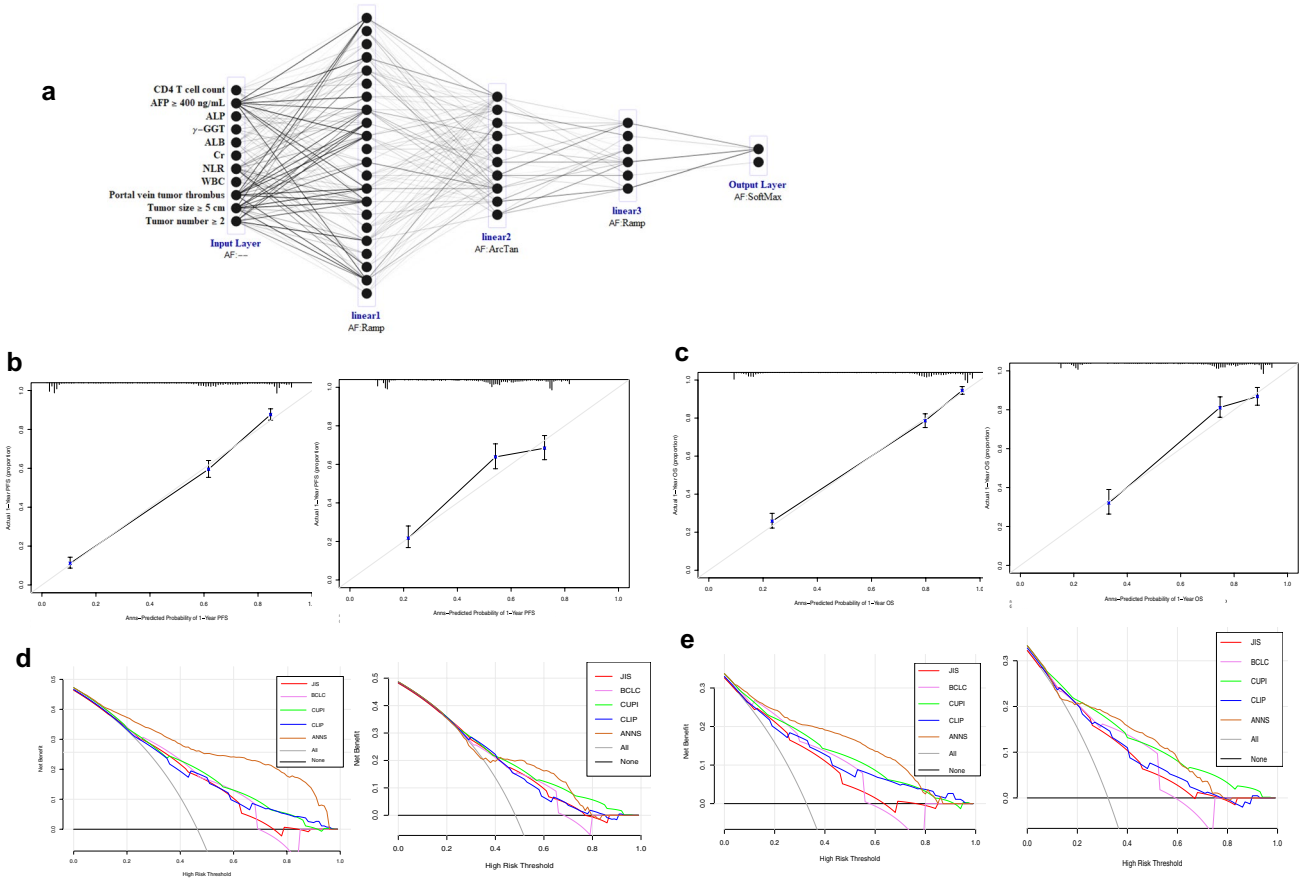

Fig. 1 ANN model, calibration plots and decision carve analysis (DCA) for predicting the outcome of HCC patients. **a** The ANNs model predicts probability of 1-year progression-free survival (PFS). Please visit https://lixuan.me/annmodel/myg-v3/. The calibration curve for predicting 1-year PFS (**b**) and 1-year OS (**c**) in training and validation cohort. The DCA for predicting 1-year PFS (**d**) and 1-year OS (**e**) in training and validation cohort

the same obvious survival differences among all stratifications (Fig. 2d, f). The ANNs model could clearly distinguish all patients according to different PFS risks, whether for the training cohort or validation cohort.

Prognostic value of CD4T cell counts in patients with HCC

We took the median value of CD4 count 445/μL as the cutoff value, and divided all patients into two groups. Through the Kaplan–Meier survival curves, it was found that the PFS and OS in the high CD4 group were significantly higher than those in the low CD4 group ($p < 0.05$) (Fig. 3a, b). However, among the high-, medium-, and low-risk subgroups differentiated by the ANNs model, it was found that the clinical benefits of high CD4 only existed in patients with low and medium risks. In Strata 3 with high risk of PFS, there was no difference in survival between patients with high CD4 and patients with low CD4 (Fig. 3c, d).

Discussion

At present, there are several staging systems used to distinguish patients with HCC, but none was specially designed to predict PFS of HCC patients. In this study, for the first time, an artificial neural network prediction model suitable for individualized patients was constructed using the method of machine learning, which can calculate the probability of the PFS of HCC within one year. It integrates the tumor characteristics of patients with HCC: Tumor size, number, portal vein tumor thrombus, AFP, liver function indicators ALB, γ-GGT, and ALP, inflammation indicators NLR and WBC, and immunity indicator CD4 T cell counts. The C indexes of the prediction model in this study were higher than 0.7.

Our results show that the ANNs model system is superior to other scoring systems in predicting the PFS of HCC. ANNs model can learn from each data and connect each input with the corresponding output by changing the weight of the connections in the neurons [23]. When compared with the logistic regression model or cox regression model in

Table 2 Comparison of the performance and discriminative ability among the current model and other models

Cohort	Models	1-year progression-free survival		1-year survival	
		AUROC (95% CI)	C-index (95% CI)	AUROC (95% CI)	C-index (95% CI)
Training	ANNs	0.866 (0.848–0.884)	0.782 (0.767–0.797)	0.877 (0.858–0.895)	0.824 (0.808–0.840)
	BCLC staging	0.739 (0.712–0.765)	0.694 (0.676–0.711)	0.790 (0.766–0.815)	0.746 (0.726–0.765)
	TNM staging	0.697 (0.669–0.725)	0.662 (0.642–0.682)	0.707 (0.678–0.736)	0.678 (0.655–0.700)
	Okuda staging	0.650 (0.621–0.679)	0.624 (0.606–0.643)	0.692 (0.662–0.722)	0.665 (0.645–0.686)
	CUPI	0.737 (0.710–0.763)	0.702 (0.683–0.721)	0.791 (0.765–0.816)	0.760 (0.739–0.781)
	CLIP	0.748 (0.722–0.773)	0.711 (0.693–0.730)	0.815 (0.791–0.840)	0.778 (0.760–0.797)
	JIS	0.745 (0.719–0.771)	0.699 (0.680–0.718)	0.778 (0.753–0.803)	0.739 (0.718–0.759)
	ALBI	0.629 (0.599–0.658)	0.604 (0.583–0.625)	0.631 (0.602–0.661)	0.636 (0.612–0.659)
	MELD	0.628 (0.6–0.657)	0.606 (0.585–0.627)	0.665 (0.636–0.695)	0.648 (0.624–0.671)
	MELD-Na	0.644 (0.616–0.672)	0.618 (0.597–0.639)	0.681 (0.652–0.71)	0.662 (0.638–0.686)
Validation	ANNs	0.730 (0.690–0.770)	0.704 (0.675–0.732)	0.804 (0.765–0.843)	0.769 (0.736–0.802)
	BCLC staging	0.706 (0.665–0.748)	0.669 (0.642–0.696)	0.706 (0.664–0.747)	0.729 (0.698–0.759)
	TNM staging	0.688 (0.646–0.731)	0.666 (0.638–0.694)	0.687 (0.645–0.730)	0.689 (0.657–0.722)
	Okuda staging	0.631 (0.587–0.676)	0.603 (0.575–0.632)	0.630 (0.586–0.675)	0.653 (0.620–0.686)
	CUPI	0.709 (0.668–0.750)	0.680 (0.650–0.709)	0.709 (0.668–0.750)	0.741 (0.708–0.774)
	CLIP	0.726 (0.686–0.750)	0.688 (0.659–0.717)	0.725 (0.685–0.766)	0.761 (0.730–0.791)
	JIS	0.726 (0.685–0.766)	0.690 (0.662–0.718)	0.724 (0.684–0.765)	0.744 (0.713–0.774)
	ALBI	0.647 (0.605–0.690)	0.615 (0.584–0.646)	0.646 (0.603–0.690)	0.657 (0.622–0.693)
	MELD	0.634 (0.592–0.677)	0.603 (0.572–0.634)	0.687 (0.644–0.73)	0.662 (0.627–0.697)
	MELD-Na	0.64 (0.598–0.683)	0.611 (0.58–0.642)	0.693 (0.65–0.736)	0.669 (0.634–0.704)

the traditional sense, ANN model is a non-linear model. It repeatedly trains the factors related to the outcome to achieve the best prediction model, so the ANN model has higher prediction accuracy. The classic TNM staging only includes tumor features, and was often applied to patients undergoing surgical resection, and showed poor prognosis prediction efficiency in advanced patients who cannot undergo surgery [24]. In addition, JIS from Japan and CUPI system from China also mainly rely on TNM, which may not well apply to predict the prognosis of all HCC population. BCLC staging has been widely used in clinical practice, but it is used in EASL and AASLD guidelines for staging patients and guiding the best treatment strategy. ALBI scores only contain two indicators, albumin and total bilirubin, which reflect liver function [22]. ALBI grading alone can hardly predict the prognosis of patients with HCC accurately, and its C-index was the lowest in this study, which was only 0.604. Okuda grading takes tumors > 50% or < 50% of the liver volume as an important judgment basis. However, with the standardization of HCC screening and the increased detection rate, most patients have a tumor volume < 50% of the liver volume; hence, Okuda is unable to further effectively distinguish patients in the early and middle stages. It is necessary to develop one model among patients with the similar tumor burdens. When Okuda was compared with other modern models, it showed lower prediction values [25]. Similar results were also found in this study. The AUROC of Okuda grading (0.650) was only higher than that of ALBI grading. The above models were also compared in a large-scale cohort study, which found that the CLIP score had the highest homogeneity and the lowest AIC value, and the same results were also found in the subgroup of HBV–HCC and the subgroup treated by radioembolization [26]. Consistent with the previous reports, our study also found that the CLIP score is second only to ANNs in predicting PFS of HCC patients, with a C-index of 0.711.

Treatment methods have definite effects on the disease progression and survival in patients with HCC. Previous studies have mostly focused on the prognosis model for a specific treatment approach [7, 8]. We also performed a subgroup analysis and found that the ANNs model could predict the PFS of patients in HCC in the resection, local treatment, or palliative treatment groups (all AUROC values > 0.8, Table S2). In addition, we also compared the prediction ability of the ANNs model in patients with different Child–Pugh grades, and obtained consistent results that the ANNs model is better than other models.

Local infiltration of immune cells in multiple solid tumors is associated with a good prognosis [27]. Some studies found that a high density of intratumoral CD3 T cells can significantly reduce the recurrence rate of HCC after resection [28]. Due to the limitation in liver tissue collection, researches on the correlation between tumor infiltrating lymphocytes (TILs) and HCC prognosis are limited to patients

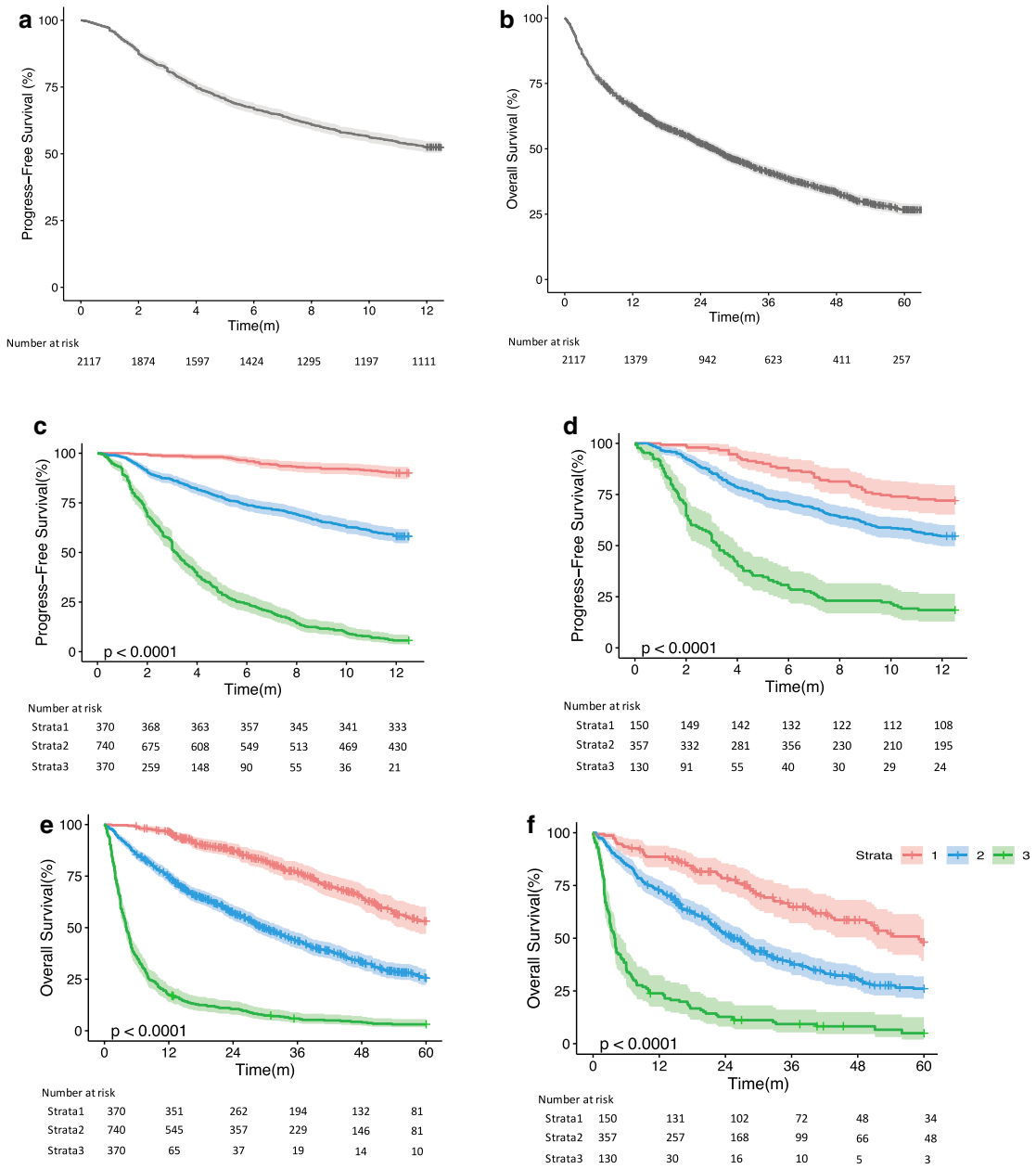

Fig. 2 The Kaplan–Meier survival curves of progression-free survival and overall survival in all HCC patients (**a**, **b**), training cohort (**c**, **e**), and validation cohort (**d**, **f**). **c**–**f** comparison of survival distributions by ANNs model

undergoing surgery or liver transplantation. However, in clinical practice, HCC is mostly found to be advanced and without hepatectomy. In this study, only 9.3% of the patients underwent surgical resection. Therefore, the influence of immune cells on the survival in patients with intermediate and advanced HCC is still unclear. This study found that a high expression of CD4 T cells could improve the PFS and OS of HCC. This is consistent with the result that the circulating T cell count was correlated with a poor prognosis in other tumor types [29]. Although CD4 cannot kill tumor cells directly, it can enhance the initiation and expansion of CD8 cells by producing IFN-γ and some chemokine [30]. On the other hand, it can recruit NK cells and type I macrophages to the tumor site, and eventually exerts a synergistic effect in tumor eradication [31]. Moreover, there is an evidence that CD4 has stronger antitumor effect than CD8, especially when the expression of MHC molecules in tumor cells is not required [32]. This provides a basis for CD4 effector cells to become an adjuvant immunotherapy in the future.

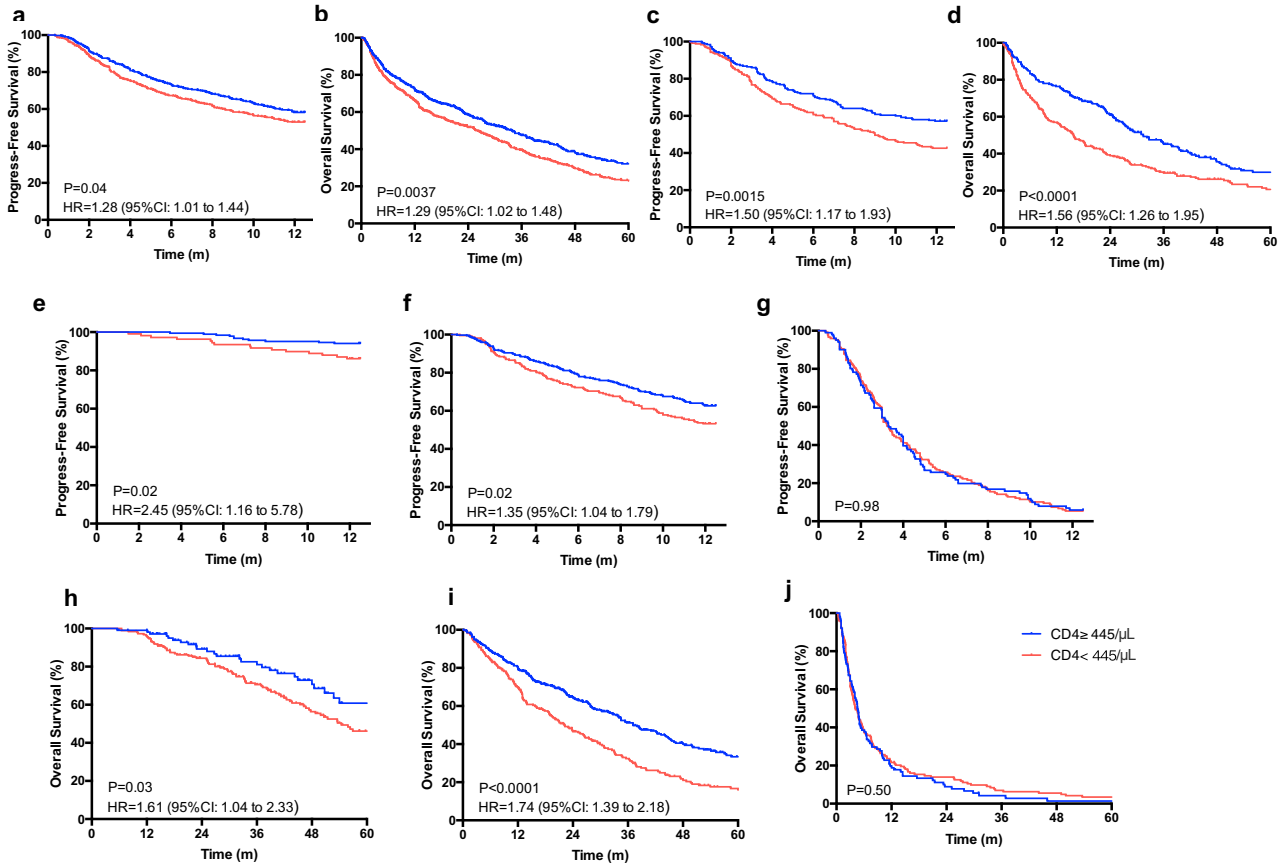

Fig. 3 The survival analysis between HCC patients with CD4 ≥ 445/μL and CD4 < 445/μL. **a**, **b** The Kaplan–Meier curves of PFS and OS in training cohort; **c**, **d** the Kaplan–Meier curves of PFS and OS in validation cohort; **e**–**g** the Kaplan–Meier curves of PFS in Strata 1, 2, 3 divided by ANNs model in training cohort; **h**–**j** the Kaplan–Meier curves of OS in Strata 1, 2, 3 divided by ANNs model in training cohort

Our research also has several limitations. First, this is a retrospective single-center study, which inevitably has selection bias. However, this study has the advantages of complete clinical data and a large sample size. Combined with deep learning and training of the artificial intelligence neural network, the prediction ability of the ANNs model for PFS greatly improved. Second, HCC is mainly related to HBV in China. Compared with HCC caused by HCV or alcoholic liver in other regions, patients with HCC in China may present different tumor features such as smaller tumor burden, etc. Therefore, it is necessary to further carry out large-scale external verification in different regions.

Summary

In this study, the artificial neural network was used to construct a prognosis model to predict the PFS of HCC. As a risk-stratification tool, the model is simple and convenient, and has good individualized prediction performance, and is helpful to evaluate PFS in patients with HCC in clinical practice. However, further verification is still needed in the future.

Funding This work was supported by the National Science Foundation of China (No. 81874435), Application of Clinical Features of Capital City of Science and Technology commission (No. Z171100001017082), Dengfeng Talent Support Program of Beijing Municipal Administration of Hospitals (No. DFL20191803).

Compliance with ethical standards

Conflict of interest The authors confirm that there are no known conflicts of interest associated with this study.

Ethical approval The study was approved by the Ethics Committee of Beijing Ditan Hospital, Capital Medical University. Written informed consent was obtained from each patient. All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2008.

References

1. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394–424.
2. Foreman KJ, Marquez N, Dolgert A, et al. Forecasting life expectancy, years of life lost, and all-cause and cause-specific mortality for 250 causes of death: reference and alternative scenarios for 2016–40 for 195 countries and territories. Lancet. 2018;392(10159):2052–2090.
3. Sapisochin G, Bruix J. Liver transplantation for hepatocellular carcinoma: outcomes and novel surgical approaches. Nat Rev Gastroenterol Hepatol. 2017;14(4):203–217.
4. He LL, Liu XL, Zhang S, et al. Independent risk factors for disease recurrence after surgery in patients with hepatitis B virus-related hepatocellular carcinoma≤ 3 cm in diameter. Gastroenterol Rep. 2019;7(4):250–257.
5. Marasco G, Colecchia A, Colli A, et al. Role of liver and spleen stiffness in predicting the recurrence of hepatocellular carcinoma after resection. J Hepatol. 2019;70(3):440–448.
6. Shiina S, Tateishi R, Arano T, et al. Radiofrequency ablation for hepatocellular carcinoma: 10-year outcome and prognostic factors. Am J Gastroenterol. 2012;107(4):569–577.
7. Shim JH, Jun MJ, Han S, et al. Prognostic nomograms for prediction of recurrence and survival after curative liver resection for hepatocellular carcinoma. Ann Surg. 2015;261(5):939–946.
8. Mehta N, Heimbach J, Harnois DM, et al. Validation of a risk estimation of tumor recurrence after transplant (RETREAT) score for hepatocellular carcinoma recurrence after liver transplant. JAMA Oncol. 2017;3(4):493–500.
9. Han K, Ren M, Wick W, et al. Progression-free survival as a surrogate endpoint for overall survival in glioblastoma: a literature-based meta-analysis from 91 trials. Neuro-Oncology. 2013;16(5):696–706.
10. Llovet JM, Montal R, Villanueva A. Randomized trials and endpoints in advanced HCC: role of PFS as a surrogate of survival. J Hepatol. 2019;70(6):1262–1277.
11. Saad ED, Buyse M. Statistical controversies in clinical research: end points other than overall survival are vital for regulatory approval of anticancer agents. Ann Oncol. 2015;27(3):373–378.
12. Bishop CM. Neural networks for pattern recognition. Oxford: Oxford University Press; 1995.
13. Amato F, López A, Peña-Méndez EM, et al. Artificial neural networks in medical diagnosis. J Appl Biomed. 2013;11(2):47–58.
14. Liu X, Li M, Wang X, et al. PD-1+ TIGIT+ CD8+ T cells are associated with pathogenesis and progression of patients with hepatitis B virus-related hepatocellular carcinoma. Cancer Immunol Immunother. 2019;68(12):2041–2054.
15. Lencioni R, Llovet JM. Modified RECIST (mRECIST) assessment for hepatocellular carcinoma. Semin Liver Dis. 2010;30(01):052–060.
16. Minagawa M, Ikai I, Matsuyama Y, et al. Staging of hepatocellular carcinoma: assessment of the Japanese TNM and AJCC/UICC TNM systems in a cohort of 13,772 patients in Japan. Ann Surg. 2007;245(6):909.
17. Forner A, Reig ME, de Lope CR, et al. Current strategy for staging and treatment: the BCLC update and future prospects. Semin Liver Dis. 2010;30(01):061–074.
18. Okuda K, Obata H, Nakajima Y, et al. Prognosis of primary hepatocellular carcinoma. Hepatology. 1984;4(S1):3S–6S.
19. Cancer of the Liver Italian Program (CLIP) Investigators. A new prognostic system for hepatocellular carcinoma: a retrospective study of 435 patients. Hepatology. 1998;28(3):751–755.
20. Leung TWT, Tang AMY, Zee B, et al. Construction of the Chinese University Prognostic Index for hepatocellular carcinoma and comparison with the TNM staging system, the Okuda staging system, and the Cancer of the Liver Italian Program staging system: a study based on 926 patients. Cancer. 2002;94(6):1760–1769.
21. Kudo M, Chung H, Osaki Y. Prognostic staging system for hepatocellular carcinoma (CLIP score): its value and limitations, and a proposal for a new staging system, the Japan Integrated Staging Score (JIS score). J Gastroenterol. 2003;38(3):207–215.
22. Pinato DJ, Sharma R, Allara E, et al. The ALBI grade provides objective hepatic reserve estimation across each BCLC stage of hepatocellular carcinoma. J Hepatol. 2017;66(2):338–346.
23. Agatonovic-Kustrin S, Beresford R. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. J Pharm Biomed Anal. 2000;22(5):717–727.
24. Huitzil-Melendez FD, Capanu M, O'Reilly EM, et al. Advanced hepatocellular carcinoma: which staging systems best predict prognosis? J Clin Oncol. 2010;28(17):2889.
25. Chen ZH, Hong YF, Lin J, et al. Validation and ranking of seven staging systems of hepatocellular carcinoma. Oncology letters. 2017;14(1):705–714.
26. Liu PH, Hsu CY, Hsia CY, et al. Prognosis of hepatocellular carcinoma: assessment of eleven staging systems. J Hepatol. 2016;64(3):601–608.
27. Fridman WH, Pagès F, Sautès-Fridman C, et al. The immune contexture in human tumours: impact on clinical outcome. Nat Rev Cancer. 2012;12(4):298.
28. Gabrielson A, Wu Y, Wang H, et al. Intratumoral CD3 and CD8 T-cell densities associated with relapse-free survival in HCC. Cancer Immunol Res. 2016;4(5):419–430.
29. Martens A, Wistuba-Hamprecht K, Yuan J, et al. Increases in absolute lymphocytes and circulating CD4+ and CD8+ T cells are associated with positive clinical outcome of melanoma patients treated with ipilimumab. Clin Cancer Res. 2016;22(19):4848–4858.
30. Phares TW, Stohlman SA, Hwang M, et al. CD4 T cells promote CD8 T cell immunity at the priming and effector site during viral encephalitis. J Virol. 2012;86(5):2416–2427.
31. Kim HJ, Cantor H. CD4 T-cell subsets and tumor immunity: the helpful and the not-so-helpful. Cancer Immunol Res. 2014;2(2):91–98.
32. Perez-Diez A, Joncker NT, Choi K, et al. CD4 cells can be more efficient at tumor rejection than CD8 cells. Blood. 2007;109(12):5346–5354.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Xiaoli Liu[1] · Yixin Hou[1] · Xinhui Wang[1] · Lihua Yu[1] · Xianbo Wang[1] · Li Jiang[2] · Zhiyun Yang[1]

[1] Center of Integrative Medicine, Beijing Ditan Hospital, Capital Medical University, No. 8 Jing Shun East Street, Beijing 100015, China

[2] Department of Hepatobiliary Surgery, Beijing Ditan Hospital, Capital Medical University, No. 8 Jing Shun East Street, Beijing 100015, China