

Machine Learning for Dynamic Prognostication of Patients With Hepatocellular Carcinoma Using Time-Series Data: Survival Path Versus Dynamic-DeepHit HCC Model

Cancer Informatics
Volume 23: 1–10
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/11769351241289719



Lujun Shen^{1,2}, Yiquan Jiang^{1,2}, Tao Zhang³, Fei Cao^{1,2}, Liangru Ke^{2,4}, Chen Li^{1,2}, Gulijayina Nuerhashi^{1,2}, Wang Li^{1,2}, Peihong Wu^{1,2}, Chaofeng Li^{2,5}, Qi Zeng⁶ and Weijun Fan^{1,2}

¹Department of Minimally invasive therapy, Sun Yat-sen University Cancer Center, Guangzhou, P. R. China. ²State Key Laboratory of Oncology in South China, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-sen University Cancer Center, Guangzhou, P. R. China. ³Department of Information, Nanfang Hospital, Southern Medical University, Guangzhou, P.R. China. ⁴Department of Radiology, Sun Yat-sen University Cancer Center, Guangzhou, P. R. China. ⁵Information Center, Sun Yat-sen University Cancer Center, Guangdong, China. ⁶Cancer center, The Fifth Affiliated Hospital of Sun Yat-sen University, Zhuhai, Guangdong, China.

ABSTRACT

OBJECTIVES: Patients with intermediate or advanced hepatocellular carcinoma (HCC) require repeated disease monitoring, prognosis assessment and treatment planning. In 2018, a novel machine learning methodology “survival path” (SP) was developed to facilitate dynamic prognosis prediction and treatment planning. One year after, a deep learning approach called Dynamic Deephit was developed. The performance of the two state-of-art models in dynamic prognostication have not been compared.

METHODS: We trained and tested the SP and Dynamic DeepHit models in a large cohort of 2511 HCC patients using time-series data. The time-series data were converted into data of time slices, with an interval of three months. The time-dependent c-index for OS at given prediction time ($t = 1, 6, 12, 18$ months) and evaluation time ($\Delta t = 3, 6, 9, 12, 18, 24, 36, 48$ months) were compared.

RESULTS: The comparison between SP model and Dynamic DeepHit-HCC model showed the latter had significant better performance at the time of initial admission. The time-dependent c-index of Dynamic DeepHit-HCC model gradually decreased with the extension of time (from 0.756 to 0.639 in the training set; from 0.787 to 0.661 in internal testing set; from 0.725 to 0.668 in multicenter testing set); while the time-dependent c-index of SP model displayed an increased trend (from 0.665 to 0.748 in the training set; from 0.608 to 0.743 in internal testing set; from 0.643 to 0.720 in multicenter testing set). When the prediction time comes to 6 months or later since initial treatment, the survival path model outperformed the dynamic DeepHit model at late evaluation times ($\Delta t > 12$ months).

CONCLUSIONS: This research highlighted the unique strengths of both models. The SP model had advantage in long term prediction while the Dynamic DeepHit-HCC model had advantages in prediction at near time points. Fine selection of models is needed in dealing with different scenarios.

KEYWORDS: Hepatocellular carcinoma, prognosis, survival path, dynamic DeepHit, time-series data

RECEIVED: May 28, 2024. **ACCEPTED:** September 11, 2024.

TYPE: Original Research Article

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported by Science and Technology Project of Guangzhou City (202201011375).

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHORS: Dr. Weijun Fan, Department of Minimally invasive therapy, Sun Yat-sen University Cancer Center, No.651, Dongfeng East Road, Yuexiu District, Guangzhou 510060, P. R. China. Email: fanwj@sysucc.org.cn.

Dr. Qi Zeng, Department of Minimally invasive therapy, Sun Yat-sen University Cancer Center, No. 651, Zhuhai, Guangzhou 519000, P. R. China. Email: zengqi37@mail.sysu.edu.cn.

Introduction

Hepatocellular carcinoma (HCC) is the sixth most common cancer and the third leading cause of cancer-related death worldwide.^{1,2} Patients with intermediate stage HCC have poor prognosis and always require dynamic disease surveillance during comprehensive treatments, which consists of interventional therapies, targeted therapies and immunotherapy.^{3,4} The reported median survival time of patients with intermediate stage HCC ranges from 16.0 to 23.9.^{5,6}

Lujun Shen, Yiquan Jiang and Tao Zhang contributed equally to this work.

The 2022 update of BCLC strategy added two new concepts, the treatment stage migration (TSM) and untreatable progression, to facilitate dynamic adjustments of treatment plan when recommended treatment is not optimal.⁷ In 2023, Alessandro Vitale proposed an evidence-based framework for treatment of HCC based on the novel concept of “multiparametric therapeutic hierarchy,” which allows dynamic adaption of the staging based algorithms.⁸ Although considerable efforts had been made for optimizing and standardizing dynamic management of HCC patients, high-quality evidence is scarce to support refined adjustment of treatment plan.⁹ Harnessing clinical big data to facilitate



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

dynamic prognosis prediction and treatment planning for HCC patients represents a substantial challenge currently.¹⁰

During the dynamic surveillance of patients with HCC, the time-series clinical data rapidly accumulate.¹¹ Modeling of these data may delineate the biological behaviors of HCC and help guide dynamic management. The classical autoregressive integrated moving average model (ARIMA) does not fit for survival data.¹² Temporal abstraction,¹³ hidden Markov models,¹⁴ and dynamic bayesian networks could be utilized for time-series survival data.¹⁵ In recent years, a deep learning model called Dynamic Deephit based on RNN and attention mechanism was proposed, which theoretically can utilize a large fraction of repeated measurements and provide predictions with high consistency.^{16,17} Although these model have high accuracy in prediction, the nature of black box models make the results difficult to interpret.

In clinical practice, an ideal model should predict prognosis precisely while also being user-friendly and can provide guidance on treatment.¹⁸ Our previous study in 2018 proposed a novel analytical approach called survival path, which converts the timeseries data into a cascading survival map, in which each survival path bifurcates at fixed time interval depending on selected prognostic features.¹¹ The model demonstrated to have higher superior or equal value than conventional staging systems in dynamic prognosis prediction for HCC patients at specific time interval. On the other hand, with the emergence of Dynamic DeepHit model in 2020, the value in dynamic prognostication between survival path model and Dynamic Deephit model for HCC patients haven't been compared.

Therefore, in this study, we set out to compare the Survival Path model and Dynamic-Deephit model for dynamically prognosticating patients with HCC. Our study yielded significant findings on multiple aspects. Firstly, feature engineering was found playing a central role in enhancing SP model's performance as well as its ability in generalization. Secondly, either the SP model or the Dynamic DeepHit model have its unique strengths in dynamic prognostication of patients with HCC. These advancements lay the groundwork for future research on developing novel machine learning tools for dynamic prognostication and management of HCC.

Methods

Study design and patient cohorts

Between January 2007 and Jan 2015, 10621 consecutive patients with newly diagnosed HCC at Sun Yat-sen University Cancer Center (SYSUCC) were retrospectively reviewed to develop the derivation (training) cohort. Between February 2015 and January 2016, an independent consecutive series of 2105 HCC patients treated at SYSUCC were reviewed to develop the internal validation cohort. Besides, between February 2016 and August 2018, 6055 patients from SYSUCC were reviewed to develop the internal testing cohort. The inclusion criteria were as follows: (1) clinically diagnosed with intermediate stage (BCLC stage B) HCC; (2) complete data of any of the following at initial diagnosis:

computed tomography (CT) or magnetic resonance imaging (MRI) of the abdominal region, radiography or CT of the chest, routine bloodwork test, biochemical routine test, serum AFP level, and coagulation indices. The exclusion criteria were: (1) with history of other malignancies; (2) ECOG PS score >1 at initial diagnosis. The Hospital Ethics Committee of SYSUCC approved this study (B2023-639-01), which waived the need for written informed consent based on the retrospective nature of the study. A total of 1000, 200, and 879 patients were included in the derivation cohort, internal validation cohort, and testing cohort, respectively. Public multi-center database of 414 patients with intermediate stage HCC from three medical centers in southern China, which contains times-series clinical data on imaging and blood tests, was utilized as multicenter testing cohort (Figure 1).

The majority of HCC patients received transarterial chemoembolization (TACE) based integrated therapies as first-line treatment, which is decided based on the decision of the multidisciplinary teams, including hepatologists, radiologists, and interventional radiologists. The subsequent therapies after failure of first-line treatment included ablation, targeted therapies and palliative chemotherapy. Patients were advised to receive followed-up monthly during the period of initial treatment, subsequently at every 2-3 months for the first 2 years if complete remission was achieved. The frequency gradually decreased to every 3-6 months after 2 years' remission.

The building of Dynamic DeepHit model was based on the derivation cohort and internal validation cohort. Survival path models were built using the derivation cohort as it does not need internal validation when setting up model.

Workflow of survival path mapping

The survival path models were built based on published R package, SurvivalPath.¹⁹ The interval for time slices was set at 3 months and survival path model with nine time slices were computed. The minimum splitting sample size was 15 and the alpha value of significance is set at 0.05. Two survival path models were built based on the training dataset: the model of raw variables and the model of curated variables. The included variables in each survival path model were list in Supplemental Table S1. The key difference between the two models is that, in the model of curated variables, the empirical binary variables with known clinical and prognostic significance were added, including variables defined by both size and number of intrahepatic lesions, variables summarize both vascular invasion and extrahepatic metastasis, and variable that describe the change of lesions. All the empirical variables can be computed based on variables that included in the model of raw variables.

Workflow of dynamic deephit model for HCC

The Dynamic Deephit model for HCC adopted the same architecture of the model described in previous work by Changhee Lee et al.¹⁶ and is developed under the Python 3.7 platform. The deep learning model contained a shared subnetwork that process

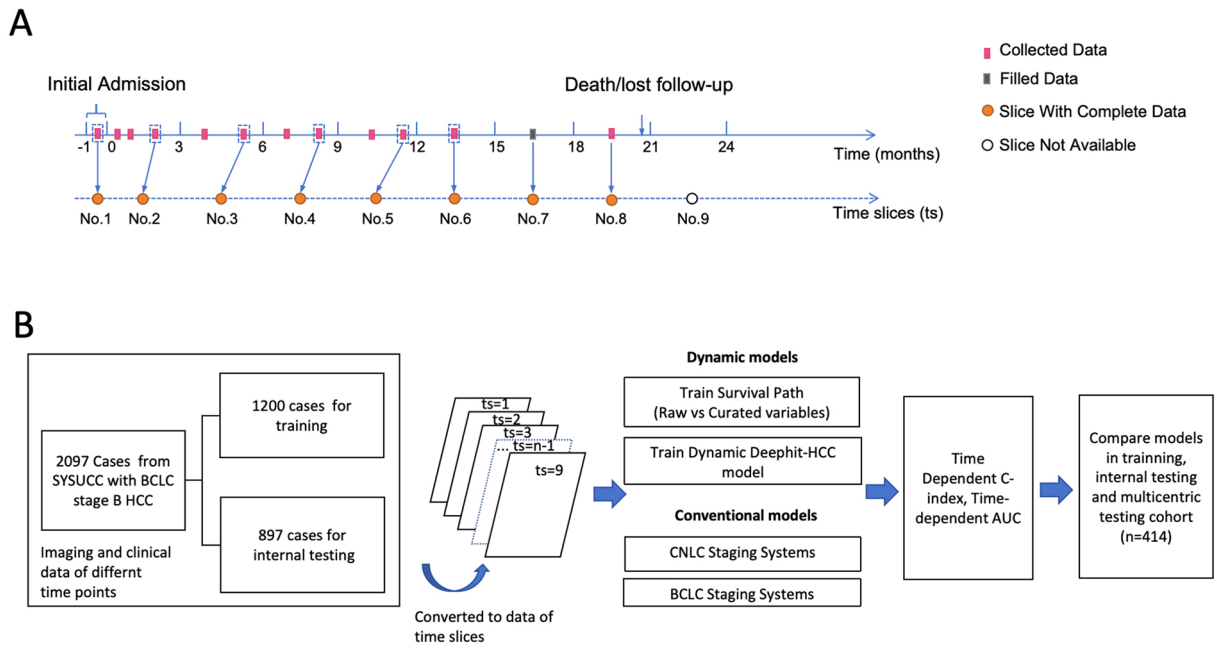


Figure 1. Flowchart of study design.

Times series data of 2097 patients with intermediate stage HCC were included in this study and were separated into training set, internal validation set and testing set, respectively: (A) The time series data were converted into data of time slices (ts) and (B) The training set was utilized to build the survival path prognostic model and internal validation set was also included when build the Dynamic DeepHit-HCC model. The time-dependent C-index of each model were compared in training set, internal testing set and multicentric testing set.

longitudinal measurements and predicts the next measurements of time-varying covariates, and a set of causespecific subnetworks which estimates the joint distribution of the first hitting time of death. Learning curves were utilized to optimize parameters of the model, including number of iteration and size of internal validation during model training. The parameters of the final model were described in Supplemental Table S2. All variables utilized in building the survival path model were included in the construction of Dynamic Deephit-HCC model. The predicted results for patients of given prediction time, denoted as t , and evaluation time, denoted as Δt , were further utilized to calculate time-dependent c-index $C(t, \Delta t)$ to assess the model's ability in dynamic prognostication.

Evaluation for prognostic significance

A total of five models were included and compared in dynamic prognostication for HCC patients, including survival path model using raw variables, survival path model with curated variables, Dynamic DeepHit-HCC model, CNLC staging system²⁰ and BCLC staging system.⁷ The comparison of prognostic significance between survival path model and other models were conducted in training dataset and testing datasets, respectively.

The Harrell's c-indexes at different time slices were utilized to assess the value of prognostication between two survival path models. The time-dependent C-index can capture dynamic performance of the regression model over time and provide more precise assessment of models' discriminative capability than Harrell's c-index. Given the unique feature of the Dynamic Deephit-HCC

model that requires specific prediction and evaluation times for making predictions, time-dependent c-index $C(t, \Delta t)$ of different models/staging systems were further compared to assess their ability in dynamic prognostication, where t indicates the prediction time which is the time when the prediction is made to incorporate dynamic predictions and Δt denotes the evaluation time which is the time elapsed since the prediction is made.

Statistical analysis

Pearson χ^2 test was used to compare categorical variables between groups, respectively. To compare the efficacy in dynamic prognosis prediction different survival path models, the measurement of c-index in each time slice was computed. To further compare the efficacy in dynamic prognosis prediction between the survival path models, dynamic deephit model and staging systems, the measurements of time-dependent c-index of specific prediction time (t) and evaluation time (Δt) were computed using pec R package; means and standard deviations were obtained via ten random sampling of two thirds of cases. For category prediction models, subgroups less than three cases were omitted to reduce inference from extreme cases when computing c-index or time-dependent c-index. A random seed was set using the base R package. The comparison of c-index and time-dependent c-index between different models was conducted using Z test method. It is estimated that 288 times of tests were conducted during the comparison of time-dependent c-index between models. To avoid false positives caused by multiple tests, a meaningful alpha value is set at 0.0001 (0.05/288). All analyses were done using R 3.6.3²¹

Table 1. Baseline characteristics of training set, internal validation set, and testing sets at initial diagnosis.

CHARACTERISTICS	TRAINING SET		INTERNAL VALIDATION SET		P-VALUE		INTERNAL TESTING SET		P-VALUE		MULTICENTER TESTING SET		P-VALUE	
	NO.	%	NO.	%	NO.	%	NO.	%	NO.	%	NO.	%	NO.	%
Age (years)					.957				.369				<.001	
<50	362	36.2	72	36.0			307	34.2			199	48.1		
≥50	638	63.8	128	64.0			590	65.8			215	51.9		
Gender					.336				.240				<.001	
Male	907	90.7	177	88.5			799	89.1			361	84.1		
Female	93	9.3	23	11.5			98	10.9			53	15.9		
HBV infection					<.001				<.001				.996	
Absent	38	3.8	21	10.5			82	9.1			15	3.6		
Present	962	96.2	179	89.5			815	90.9			399	96.4		
AFP (IU/ml)					.140				.773				.365	
<25	326	32.6	76	38.0			298	33.2			146	35.3		
≥25	674	67.4	124	62.0			599	66.8			268	64.7		
Child Pugh Class					.101				<.001				.264	
A	868	86.8	182	91.0			848	94.5			360	87.0		
B	132	13.2	18	9.0			49	5.5			54	13.0		
Tumor Size (cm)					.129				.255				.900	
<5	333	33.3	55	27.5			322	35.9			140	33.8		
≥5	667	66.7	145	72.5			575	64.1			274	66.2		
Number of lesions					.597				<.001				.882	
<4	395	39.5	75	37.5			285	31.8			166	40.1		
≥4	605	60.5	125	62.5			612	68.2			248	59.9		

Note. The distributions of baseline characteristics between training set and validation/testing sets are compared based on Chi-square test.

Results

The baseline characteristic of the patients

The training set consisted of 1000 HCC patients with a median age of 55 years (range, 14-85), the internal validation set consists of 200 HCC patients with a median age of 55 years (range, 20-82) and the testing set consisted of 897 HCC patients with a median age of 56 (range, 18-88) years (Table 1). The median follow-up time were 25.5, 22.3, and 24.6 months for the training set, internal validation set and internal testing set, respectively. Compared to the training set, the internal validation set and testing set had a lower proportion of patients with HBV infection. Besides, the proportion of patients with Child Pugh A class and those with multiple intrahepatic lesions (≥4) in testing set were higher compared to the training set. Compared to the training set, the multicentric testing

set had a higher proportion of patients with young age and female gender.

The survival paths built by training set with two strategies of feature selection

Although it's widely accepted feature engineering is of vital importance for prognosis modeling, it's impact in survival path mapping dealing with time-series data remains unknown. Two sets of variables, one consists of only raw variables and one consists of both raw and curated variables, were design to compute the survival path models. In the survival path model with curated variables, binary variables in terms of tumor diameter,²² number of intrahepatic tumors, liver function,²³ AFP level,^{24,25} change of lesions after treatment and composite variable were designed, and added.²⁶

Table 2. The variables included in building the survival path models and Dynamic DeepHit-HCC model.

VARIABLES	SURVIVAL PATH (RAW VARIABLES)	SURVIVAL PATH (CURATED VARIABLES)	DYNAMIC DEEPHIT-HCC
Age	Continuous variable	Continuous variable	Continuous variable
Laboratory tests			
Serum AFP level	Continuous variable	Continuous variable, <200 vs ≥200; <400 vs ≥400	Continuous variable
Child Pugh class	Ordinal variable (1,2,3)	Ordinal variable, Class B/C vs class A; class C vs class A/B	/
ALB level	/	/	Continuous variable
TBIL level	/	/	Continuous variable
PT	/	/	Continuous variable
HBV infection	/	/	Present/Absent
Imaging examination			
Diameter of largest intrahepatic lesion (mm)	Continuous variable	Continuous variable; ≤50 vs >50; ≤70 vs >70; ≤100 vs >100	Continuous variable
Number of hepatic lesions	Continuous variable	Continuous variable, <4 lesions vs ≥4 lesions	Continuous variable
Vascular invasion	With vs without	With vs without	With vs. without
Distant metastasis	With vs without	With vs without	With vs. without
Lymph Node metastasis	With vs without	With vs without	With vs. without
Lymph Node or distant metastasis/ Vascular Invasion (NMV)	/	With vs without	/
Number and size of hepatic lesions	/	≤1 lesion/2–3 lesions, D ≤30 mm vs >3 lesions/2–3 lesions, D > 30 mm	/
Number and size of hepatic lesions, NMV	/	≤1 lesion/2–3 lesions, D ≤30 mm, without NMV vs >3 lesions/2–3 lesions, D > 30 mm/ NMV	/
Change of lesions	/	With viable lesion vs. without viable lesion; With new lesion/ without new lesion	With new lesion/ without new lesion
Ascites	/	/	Present/Absent; Massive/ not Massive

Note. The source data in building survival path models and Dynamic DeepHit-HCC were the same. All variables in building survival path models can be computed by the data of building Dynamic DeepHit-HCC model.

Compared to the eight variables in the SP model of raw variables, a total of twenty variables were put into the training of the SP model of curated variables (Table 2). The time-series data of HCC patients were firstly divided into data of time slices, with an interval of 3 months, and the two SP models were displayed in Figure 2. The included variables and their cutoff of each models were listed in Supplemental Table S1. In both training set and testing set, the c-index of the two models underwent rapidly rise from time slice no. 1 to time slice no. 3 and the trend become gentle in the following time slices. Except for time slice No. 2, SP model of curated variables have superior or non-inferior performance compared to the SP model of raw variables at all other time slices (Figure 3A). The

peak c-index for both models were 0.841 and 0.786 in the training set, 0.761 and 0.722 in the internal testing set and 0.784 and 0.718 in the multicentric testing set, respectively.

Comparison of the survival path model with Dynamic DeepHit-HCC model and conventional staging systems

The value of prognostic prediction between the survival path model, Dynamic Deephit-HCC model and conventional staging system were compared at different prediction times in both training and testing datasets (Table 3). It's interesting to note that Dynamic DeepHit-HCC had significantly better

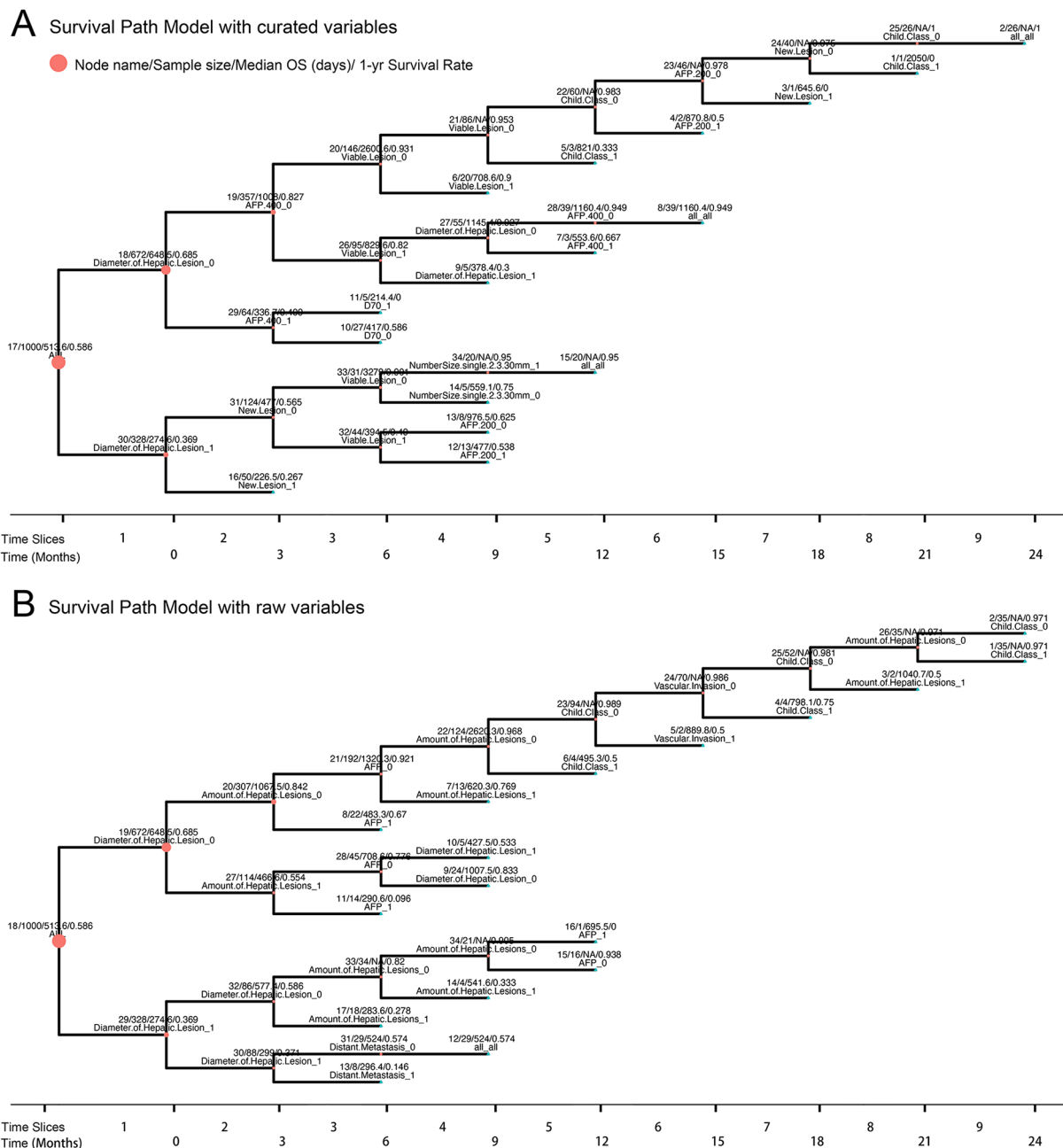


Figure 2. The survival path models constructed based on training set.

The time series data were trained on two different set of variables: (A) This plot represent survival path model built on curated variables; the variables include raw variables as well as curated variables based on previous literature and (B) This plot represent survival path model built on raw variables. When no bifurcation variable identified, the path stay the same in the following time slices.

performance in prognosis prediction compared to survival path model for patients initially admitted (prediction time at 1 month). However, the time-dependent c-index of Dynamic DeepHit-HCC model gradually decreased with the extension of evaluation time. By contrast, the fluctuation of time-dependent c-index of survival path model was relatively stable. When the prediction time is more than 6 months since onset of the treatment, the survival path model significantly outperforms the dynamic DeepHit model in long-term prognosis prediction (evaluation time > 12 months) (Figure 3B-D). The predictive ability of survival path models at different time points is

superior or not inferior to traditional CNLC and BCLC staging systems (Supplemental Table S3). The feature importance of key variables at critical time points in Dynamic Deephit-HCC model were displayed in Table 4. The influential variables of Survival Path model at corresponding time slice were displayed in Table 5. It is notable that the variable “New lesion” appeared two times in survival path model by time slice no. 7 while in Dynamic Deephit-HCC model the importance of the variable rank 9th to 10th among the 15 variables included. Besides, the increased risk of “D. of Hep Lesions” become negative at distant prediction time ($t=18$; $\Delta t=18$).

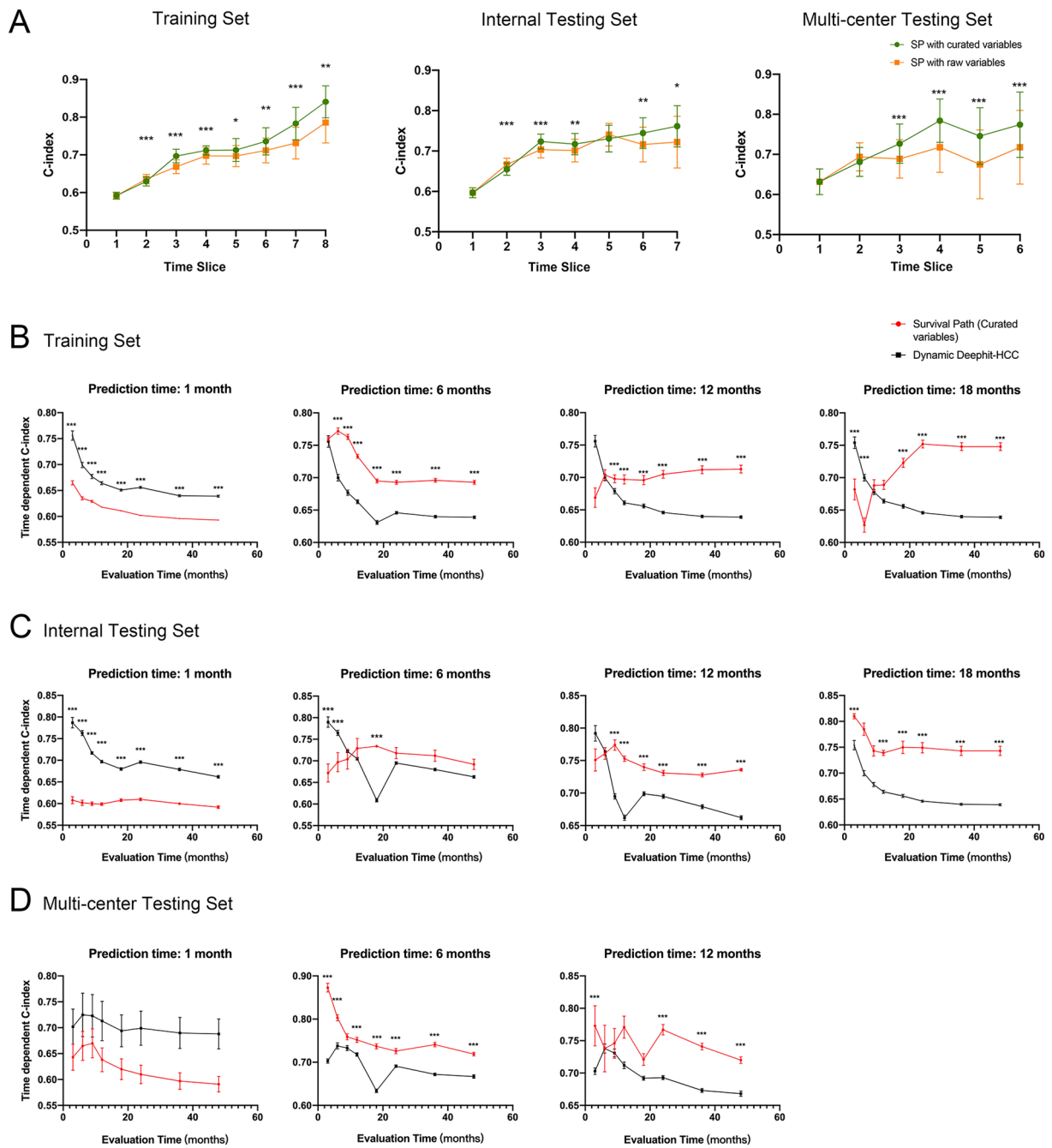


Figure 3. Comparison of C-index and time dependent c-index between different models: (A) The line chart represent change of c-index of survival path models along with time slices. The c-index of the two models underwent rapidly rise from time slice no. 1 to time slice no. 3 and the trend become gentle in the following time slices. Except for time slice No. 2, Survival path model built on curated variables have superior or non-inferior performance compared to the model built on raw variables at different time slices. The change of time dependent c-index along with different evaluation time for the two machine learning models in the training set (B), internal testing set (C) and multicentric testing set (D). ****P*-value < .0001.

Discussion

To our knowledge, this is the first study to compare the performance of two state-of-the-art machine learning approaches (Survival Path and Dynamic-Deephit) for dynamic prognosis prediction in HCC patients. Our study found that the Dynamic DeepHit-HCC model has higher predictive ability for patient at initial admission. However, when the prediction time is 6 months or even later since onset of treatment, the survival path model significantly outperforms the dynamic DeepHit

model in long-term prognosis prediction. These results suggested that although deep learning models have significant advantages in high-throughput data and accuracy in prediction, the tree based models such as survival path models also have unique advantages in dynamic prediction based on time series data when key features identified. In addition, by contrast to the nature of black box model of deep learning, the survival path model has better visualization ability and is therefore may easier be accepted by oncologists.

Table 3. The top 10 most influential covariates of dynamic Deephit-HCC model at specific prediction time.

RANK	PREDICTION (t) & EVALUATION (Δt) TIME			
	$t=1; \Delta t=3$	$t=6; \Delta t=6$	$t=12; \Delta t=6$	$t=18; \Delta t=18$
1	D. of Hep Lesions (+0.00065)	D. of Hep Lesions (+0.00120)	D. of Hep Lesions (+0.001140)	D. of Hep Lesions (-0.03951)
2	Amount of Hep Lesions (+0.00052)	ALB (-0.00097)	Amount of Hep Lesions (+0.00092)	TBLT (-0.02733)
3	ALB (-0.00051)	Amount of Hep Lesions (+0.00096)	ALB (-0.00091)	ALB (+0.01988)
4	TBLT (+0.00031)	TBLT (+0.00059)	TBLT (+0.00055)	Amount of Hep Lesions (-0.00738)
5	HBV infection (+0.00020)	HBV infection (+0.00038)	HBV infection (+0.00036)	PT (-0.00350)
6	AFP (+0.00014)	AFP (+0.00027)	AFP (+0.00026)	Age (-0.00240)
7	PT (+0.00012)	PT (+0.00022)	PT (+0.00020)	HBV infection (-0.00204)
8	Age (+0.00007)	Age (+0.00014)	Age (+0.00013)	Vascular Invasion (+0.00184)
9	New Lesion (-0.00006)	New Lesion (-0.00011)	New Lesion (-0.00011)	AFP (-0.00178)
10	Vascular Invasion (-0.00005)	Vascular Invasion (-0.00009)	Vascular Invasion (-0.00009)	New Lesion (+0.00168)

Note. The values indicate the amount of increase(+)/decrease(-) in the predicted risks on average and the covariates are ranked by the absolute values.

Table 4. The influential variables of Survival Path model using curated variables at specific prediction.

VARIABLES	PREDICTION (T , MONTHS) & CORRESPONDING TIME SLICE			
	$t=1$ (TIME SLICE=1)	$t=6$ (TIME SLICE=3)	$t=12$ (TIME SLICE=5)	$t=18$ (TIME SLICE=7)
1	D. of Hep Lesions (times=1)	D. of Hep Lesions (times=1)	D. of Hep Lesions (times=2)	D. of Hep Lesions (times=2)
2		AFP 400 (times=1)	AFP 400 (times=2)	AFP 400 (times=2)
3		New Lesion (times=1)	New Lesion (times=1)	New Lesion (times=2)
4		Viable Lesion (times=2)	Viable Lesion (times=2)	Viable Lesion (times=2)
5		D70 (times=1)	D70 (times=1)	D70 (times=1)
6			≤ 1 lesion/2–3 lesions, $D \leq 30$ mm vs >3 lesions/2–3 lesions, $D > 30$ mm (times=1)	≤ 1 lesion/2–3 lesions, $D \leq 30$ mm vs >3 lesions/2–3 lesions, $D > 30$ mm (times=1)
7			AFP 200 (times=1)	AFP 200 (times=2)
8			Child-Pugh class (times=1)	Child-Pugh class (times=1)

Note. The key bifurcation variables were same across different evaluation time given specific prediction time.

In recent years, the application of traditional machine learning models in predicting prognosis, including support vector machines, Bayesian models, and linear regression, is gradually decreasing with the development of deep learning technologies.²⁷ Tree models still have certain advantages over deep learning models because of their nature of white box model and have no strict restriction on sample size.²⁸ Numerous studies have shown that the selection of key features in tree model is crucial for its effectiveness.²⁹ Although the original data used in the SP model with curated variables and SP with raw

variables are the same, we have added variables with classical cutoff values in SP with curated variables, including the widely accepted serum AFP cutoff values and tumor diameter cutoff values. In addition, we have also added composite variables used in BCLC staging³⁰ and CNLC staging,³¹ and variables for describing tumor changes based on time series.³² In SP with curated variables, a total of 12 new variables were added, and we found that these variables can significantly enhance the model's ability to dynamically predict prognosis and its generalization ability in the testing sets. Our result suggests that a

well trained survival path model suitable for dynamic prognosis prediction of cancer requires identification and inclusion of key classification features.

To optimize the the Dynamic DeepHit-HCC model, learning curves were utilized for parameter setting the key model parameters. The Dynamic DeepHit-HCC model achieved good performance in both the training and testing sets. Compared to the Survival Path model, a peerless advantage of the Dynamic Deephit-HCC model is its ability to predict the survival probability of any evaluation time from any prediction time. In addition, at the initial admission of HCC patients, the Dynamic DeepHit-HCC model has significantly better predictive accuracy for prognosis than the Survival Path model and conventional staging systems. The reason for this phenomenon may be that Dynamic DeepHit, as a deep learning model, can directly integrate effective information of all variables for prediction since initial admission,³³ while survival paths bifurcated using only one key variable (in this case, the diameter of intrahepatic lesion) at time slice no. 1, with information in other variables unexploited. The disadvantage of Dynamic Deephit-HCC is that the model's long terms predictive ability (evaluation time > 12 months) gradually decreases. The reason for this trend may be because the majority of data in training the model are of early time slices, and hence the model has abundant raw data for short-term prediction, while in terms of long-term prediction, the fact of limited data may impede the performance of deep learning models.³⁴ In contrast, the bifurcation of the Survival Path model depends on the selection of key variables at at each time slices, and its long term prognostic ability is highly correlated with key factor identified by the data of distant time slice. Besides, the variables used in the survival path model are key prognostic factors that are fundamentally recognized in clinical practice, thereby minimizing the risk of overfitting. These two key differences in building models may explain the phenomenon that Survival Path model has better long term predictive ability for HCC patients. The differences in prognostic ability between the SP model and the Dynamic Deephit model-HCC suggested they can be applied in different scenarios to facilitate personalized prognosis prediction.

Compared to the Dynamic DeepHit model, one major drawback of the Survival Path model is that the requirement of sample size increase exponentially with the increase of time slices. If the sample size in specific node is not large enough to bifurcate when there do exist key variable that predict prognosis, the associated information will be lost. In this study, due to the restriction of sample size, we did not conduct further analysis after the ninth time slice. Developing node fusion technique to combine similar nodes may overcome this dilemma as the sample sizes of these node could be added up to support further bifurcation. In addition, the predictive ability of survival pathway model for patients at initial admission is relatively week, partly due to insufficient utilization of initial data and limited information carried by

single categorical variables. Multi-classification of the nodes at early time slices may improve the performance of the survival path model in predicting prognosis. There still remains a lot of exploratory work to be done in improving the Survival Path model.

The interpretability of machine learning models is particularly important in the clinical medicine.³⁵ Although the survival path model could be visualized using tree diagram, its interpretation is still complex. In recent years, increasing studies have shown that large language models (LLM) can facilitate interpretation of machine learning models.³⁶ Establishing an online LLM platform to enable more convenient invocation and interpretation of survival path model might have a positive impact on the promotion and application of this methodology.

Our study has several limitations. First, it is a retrospective study using dataset of single center. Second, the time-series data after patients lost surveillance/follow-up could not be obtained and hence were unable to be included in the analysis, which may be a source of bias. Therefore, to demonstrate the value of this methodology in HCC, a large-scale multi-institutional prospective study for both modeling and validation is needed.

Conclusions

Dynamic DeepHit-HCC model had advantages over survival path model in dynamic prognostication of HCC patients at early prediction times. The survival path model outperforms the dynamic DeepHit model in long-term prognosis prediction (evaluation time > 12 months) at late prediction times (≥ 6 months). The SP models, given its features of easy to visualize, use and understand, have the potential to enter the clinic in the near future.

Acknowledgements

We would like to thank Ms. Juan Nie for providing continuous encouragement to Dr. Lujun Shen in the past seven years in the pursuit of excellence in medicine. We would also like to thank Mr. Tongzhou Wang, Dr. Mengxuan Cui and Dr. Yao Wang for technical support in programming.

Authors' Contributions

Lujun Shen and Weijun Fan and conceptually designed the research; YJ, TZ, FC and LS collected data and screened data; LS and YJ analyzed the data; LS, YJ, TZ, FC, LK, CL, GN, WL, PW, CL, QZ and WF validate the interpretation and drafted and revised the manuscript. All authors have read and approved the final manuscript.

Availability of data and materials

The authenticity of this data has been validated by uploading the critical raw data onto the Research Data Deposit public platform (www.researchdata.org.cn).

ORCID iD

Weijun Fan  <https://orcid.org/0000-0002-8757-647X>

Supplemental Material

Supplemental material for this article is available online.

REFERENCES

- Konyon P, Ahmed A, Kim D. Current epidemiology in hepatocellular carcinoma. *Expert Review Gastroenterol Hepatol.* 2021;15(11):1295-1307. doi:10.1080/17474124.2021.1991792
- Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2024;74(3):229-263. doi:10.3322/caac.21834
- Brown ZJ, Tsilimigras DI, Ruff SM, et al. Management of hepatocellular carcinoma: a review. *JAMA Surg.* 2023;158(4):410-420. doi:10.1001/jamasurg.2022.7989
- Elshaarawy O, Gomaa A, Omar H, Rewisha E, Waked I. Intermediate stage hepatocellular carcinoma: a summary review. *J Hepatocell Carcinoma.* 2019;6:105-117. doi:10.2147/JHC.S168682
- Liu W, Wang K, Bao Q, Sun Y, Xing BC. Hepatic resection provided long-term survival for patients with intermediate and advanced-stage resectable hepatocellular carcinoma. *World J Surg Oncol.* 2016;14:62. doi:10.1186/s12957-016-0811-y
- Zhang W, Tong S, Hu B, et al. Lenvatinib plus anti-PD-1 antibodies as conversion therapy for patients with unresectable intermediate-advanced hepatocellular carcinoma: a single-arm, phase II trial. *J Immunother Cancer.* 2023;11(9) doi:10.1136/jitc-2023-007366
- Reig M, Forner A, Rimola J, et al. BCLC strategy for prognosis prediction and treatment recommendation: the 2022 update. *J Hepatol.* 2022;76(3):681-693. doi:10.1016/j.jhep.2021.11.018
- Vitale A, Cabibbo G, Iavarone M, et al. Personalised management of patients with hepatocellular carcinoma: a multiparametric therapeutic hierarchy concept. *Lancet Oncol.* 2023;24(7):e312-e322. doi:10.1016/S1470-2045(23)00186-9
- Lencioni R, Chen XP, Dagher L, Venook AP. Treatment of intermediate/advanced hepatocellular carcinoma in the clinic: how can outcomes be improved? *Review. Oncologist.* 2010;15 Suppl 4:42-52. doi:10.1634/theoncologist.2010-S4-42
- Chen B, Garmire L, Calvisi DF, Chua MS, Kelley RK, Chen X. Harnessing big "omics" data and AI for drug discovery in hepatocellular carcinoma. *Nat Rev Gastroenterol Hepatol.* 2020;17(4):238-251. doi:10.1038/s41575-019-0240-9
- Shen L, Zeng Q, Guo P, et al. Dynamically prognosticating patients with hepatocellular carcinoma through survival paths mapping based on time-series data. *Nat Commun.* 2018;9(1):2230. doi:10.1038/s41467-018-04633-7
- Beard E, Marsden J, Brown J, et al. Understanding and using time series analyses in addiction research. *Addiction.* 2019;114(10):1866-1884. doi:10.1111/add.14643
- Batal I, Sacchi L, Bellazzi R, Hauskrecht M. A temporal abstraction framework for classifying clinical temporal data. *AMIA Annual Symposium Proceedings AMIA Symposium.* 2009;2009:29-33.
- Chadeau-Hyam M, Tubert-Bitter P, Guihenneuc-Jouyaux C, et al. Dynamics of the risk of smoking-induced lung cancer: a compartmental hidden Markov model for longitudinal analysis. *Epidemiology.* 2014;25(1):28-34. doi:10.1097/EDE.0000000000000032
- Orphanou K, Stassopoulou A, Keravnou E. Temporal abstraction and temporal Bayesian networks in clinical domains: a survey. *Artif Intell Med.* 2014;60(3):133-49. doi:10.1016/j.artmed.2013.12.007
- Lee C, Yoon J, Schaar MV. Dynamic-DeepHit: a deep learning approach for dynamic survival analysis with competing risks based on longitudinal data. *IEEE Trans Biomed Eng.* 2020;67(1):122-133. doi:10.1109/TBME.2019.2909027
- Lee C, Light A, Saveliev ES, van der Schaar M, Gnanapragasam VJ. Developing machine learning algorithms for dynamic estimation of progression during active surveillance for prostate cancer. *NPJ Digit Med.* 2022;5(1):110. doi:10.1038/s41746-022-00659-w
- He M, Cao Y, Chi C, et al. Research progress on deep learning in magnetic resonance imaging-based diagnosis and treatment of prostate cancer: a review on the current status and perspectives. *Front Oncol.* 2023;13:1189370. doi:10.3389/fonc.2023.1189370
- Shen L, Mo J, Yang C, et al. SurvivalPath: A R package for conducting personalized survival path mapping based on time-series survival data. *PLoS Comput Biol.* 2023;19(1):e1010830. doi:10.1371/journal.pcbi.1010830
- Xie DY, Ren ZG, Zhou J, Fan J, Gao Q. 2019 Chinese clinical guidelines for the management of hepatocellular carcinoma: updates and insights. *Hepatobiliary Surg Nutr.* 2020;9(4):452-463. doi:10.21037/hbsn-20-480
- R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria 2020. <https://www.R-project.org/>
- Benson AB, D'Angelica MI, Abbott DE, et al. Hepatobiliary cancers, Version 2.2021, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw.* 2021;19(5):541-565. doi:10.6004/jnccn.2021.0022
- Wang Z, Fan Q, Wang M, et al. Comparison between Child-Pugh Score and albumin-bilirubin grade in patients treated with the combination therapy of transarterial chemoembolization and sorafenib for hepatocellular carcinoma. *Ann Transl Med.* 2020;8(8):537. doi:10.21037/atm.2020.02.114
- Lim DH, Casadei-Gardini A, Lee MA, et al. Prognostic implication of serum AFP in patients with hepatocellular carcinoma treated with regorafenib. *Future Oncol.* 2022;18(27):3021-3030. doi:10.2217/fon-2022-0524
- Hsu CY, Liu PH, Lee YH, et al. Using serum alpha-fetoprotein for prognostic prediction in patients with hepatocellular carcinoma: what is the most optimal cutoff? *PLoS One.* 2015;10(3):e0118825. doi:10.1371/journal.pone.0118825
- Lencioni R, Llovet JM. Modified RECIST (mRECIST) assessment for hepatocellular carcinoma. *Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Review. Sem Liver Dis.* 2010;30(1):52-60. doi:10.1055/s-0030-1247132
- Albaradei S, Thafar M, Alsaedi A, et al. Machine learning and deep learning methods that use omics data for metastasis prediction. *Comput Struct Biotechnol J.* 2021;19:5008-5018. doi:10.1016/j.csbj.2021.09.001
- Nayariseri A, Khandelwal R, Tanwar P, et al. Artificial intelligence, big data and machine learning approaches in precision medicine & drug discovery. *Curr Drug Targets.* 2021;22(6):631-655. doi:10.2174/1389450122999210104205732
- Kamkar I, Gupta SK, Phung D, Venkatesh S. Stable feature selection for clinical prediction: exploiting ICD tree structure using Tree-Lasso. *J Biomed Inform.* 2015;53:277-90. doi:10.1016/j.jbi.2014.11.013
- Duseja A. Staging of hepatocellular carcinoma. *J Clin Exp Hepatol.* 2014;4(Suppl 3):S74-9. doi:10.1016/j.jceh.2014.03.045
- Xie DY, Zhu K, Ren ZG, Zhou J, Fan J, Gao Q. A review of 2022 Chinese clinical guidelines on the management of hepatocellular carcinoma: updates and insights. *Hepatobiliary Surg Nutr.* 2023;12(2):216-228. doi:10.21037/hbsn-22-469
- Hwang YJ, Lee Y, Park H, Lee Y, Lee K, Kim H. Prognostic significance of viable tumor size measurement in hepatocellular carcinomas after preoperative locoregional treatment. *J Pathol Transl Med.* 2021;55(5):338-348. doi:10.4132/jptm.2021.07.26
- Huang S, Yang J, Fong S, Zhao Q. Artificial intelligence in cancer diagnosis and prognosis: Opportunities and challenges. *Cancer letters.* Feb 28 2020;471:61-71. doi:10.1016/j.canlet.2019.12.007
- Tran KA, Kondrashova O, Bradley A, Williams ED, Pearson JV, Waddell N. Deep learning in cancer diagnosis, prognosis and treatment selection. *Genome Med.* 2021;13(1):152. doi:10.1186/s13073-021-00968-x
- Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. *Lancet Oncol.* 2019;20(5):e262-e273. doi:10.1016/S1470-2045(19)30149-4
- Benary M, Wang XD, Schmidt M, et al. Leveraging large language models for decision support in personalized oncology. *JAMA Netw Open.* 2023; 6(11):e2343689. doi:10.1001/jamanetworkopen.2023.43689