

A machine learning algorithm facilitates treatment selection for up to three hepatocellular carcinomas ≤ 3 cm in size

Takashi Kokudo

Japan Institute for Health Security, National Center for Global Health and Medicine

Yasuhide Yamada

National Center for Global Health and Medicine

Yoshinari Asaoka

Teikyo University School of Medicine

Ryosuke Tateishi

The University of Tokyo

Kiyoshi Hasegawa

The University of Tokyo

Yoshinori Kabeya

Healthcare & Life Sciences, IBM Japan

Ami Kamada

Healthcare & Life Sciences, IBM Japan

Sumito Yoshida

Japan Medical Association Research Institute

Yuma Nakamura

Healthcare & Life Sciences, IBM Japan

Kengo Yoshimitsu

Fukuoka University

Hirohisa Yano

Kurume University

Takamichi Murakami

Kobe University

Takumi Fukumoto

Kobe University

Etsuro Hatano

Kyoto University

Mitsuo Shimada

Tokushima University

Naoya Kato

Chiba University

Hiroko Iijima

Hyogo Medical University

Masayuki Kurosaki

Musashino Red Cross Hospital

Michiie Sakamoto

International University of Health and Welfare

Masatoshi Kudo

Kindai University

Norihiro Kokudo

kokudo.n@jihs.go.jp

Japan Institute for Health Security, National Center for Global Health and Medicine

Article

Keywords: hepatocellular carcinoma, radiofrequency ablation, liver resection, deep learning

Posted Date: May 25th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9218610/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Liver resection (LR) and radiofrequency ablation (RFA) are recommended for patients with up to three hepatocellular carcinomas (HCCs) measuring ≤ 3 cm and preserved liver function. Nevertheless, treatment allocation is decided according to the preference of the treating doctor. This study aimed to develop a predictive model to guide treatment decisions based on survival outcomes.

Methods

A total of 18,958 patients with HCC with up to three nodules measuring ≤ 3 cm were included from the nationwide survey of Japan. Deep survival analysis was performed using the Recurrent Deep Survival Machines (RDSM) model. Model performance was evaluated using 10-fold cross-validation, concordance index (C-index), and overall survival (OS). Survival curves were compared using the log-rank test. Potential confounding factors were identified using 1:1 propensity score matching (PSM).

Results

Patients who underwent LR showed significantly longer OS than those who underwent RFA (5-year survival rate 81.4% vs. 73.1%; $P < 0.005$). The trained RDSM model achieved a C-index of 0.68. In the deep learning (DL) model, patients who received the recommended treatment exhibited significantly longer survival than those who did not (5-year survival rate 81.2% vs. 73.9%; $P < 0.005$; PSM, 81.9% vs. 76.7%; $P < 0.005$). The DL model recommended LR in 6,966 (84.5%) patients who received RFA, especially those exhibiting typical imaging patterns (early enhancement and washout in the computed tomography (CT) images [85.9% vs. 61.7% and 84.1 vs. 74.3%, respectively]).

Conclusions

DL modeling effectively helped treatment allocation for patients with HCC with up to three nodules measuring ≤ 3 cm. Our study indicates the potential utilization of DL modeling in the treatment allocation of these patients.

Introduction

Primary liver cancer is the sixth most common cancer and the third leading cause of cancer death globally.¹ Hepatocellular carcinoma (HCC), accounting for 80%–85% of primary liver cancer cases, usually develops in patients with chronic liver diseases, including chronic hepatitis B or C infection, alcohol-associated liver disease, and metabolic dysfunction–associated liver disease. All HCC treatment guidelines recommend that clinical management decisions consider both the patient's liver function and tumor characteristics.^{2–4} However, the indications for specific treatment modalities vary among countries and regions. For instance, the European and American treatment guidelines recommend liver

resection (LR) for patients with a single nodule without portal hypertension (HT), whereas the Asian treatment guidelines adopt broader criteria to include patients with up to three nodules.

LR and radiofrequency ablation (RFA) are potentially curative treatment modalities for patients with early-stage HCC. Several randomized controlled trials comparing RFA and LR indicate that both treatments generated comparable survival outcomes for patients with 3 or fewer HCC nodules measuring ≤ 3 cm and preserved liver function.⁵⁻⁸ However, this category encompasses a heterogeneous group of patients. For instance, among patients with nodules measuring ≤ 3 cm, local tumor control rates after ablation tend to decline as tumor diameter increases. Further, patients with Child–Pugh class A liver function range from those with nearly normal hepatic reserve to those at risk of postoperative decompensation. Therefore, treatment strategies should be tailored to each patient according to detailed clinical information.

In recent years, the rapid advancement of artificial intelligence (AI), particularly in deep learning (DL)⁹ has extended to the field of survival analysis, which has traditionally been employed to assess treatment and medication effectiveness, leading to the emergence of deep survival analysis,¹⁰ including DeepSurv.¹¹ DeepSurv is developed by replacing certain linear components of the hazard function with a DL-based model. Since the introduction of DeepSurv, various extensions have been developed, including Deep Survival Machines (DSM)¹², which replace the entire survival curve estimation with a DL approach. DSM is a DL-based survival analysis model that learns flexible parametric distributions for time-to-event prediction. Unlike traditional Cox proportional hazards models, DSM does not rely on the proportional hazards assumption; instead, it directly models survival functions using a mixture of parametric distributions. The Cox proportional hazards model assumes linear covariate effects. Therefore, when complex interactions exist—such as treatment effects varying by patient characteristics and/or tumor stage—the interactions must be explicitly prespecified in the model. However, comprehensively capturing such nonlinear patterns is challenging. In contrast, the DSM approach facilitates flexible modeling of complex survival patterns without requiring explicitly defined nonlinear relationships or assuming the proportional hazards condition.

Recurrent DSM (RDSM) is a natural fit to model problems where there are time-dependent covariates. RDSM¹³ builds on the original DSM model and enables learning of representations of the input covariates using recurrent neural networks, including Long Short Term Memory¹⁴ and Gated Recurrent Unit.¹⁵ Therefore DL-based approach may provide significant advantages in predicting survival rate than traditional survival models. We developed a DL model for guiding curative treatments for patients with HCCs with up to three nodules measuring ≤ 3 cm based on survival using the nationwide survey data from the Liver Cancer Study Group of Japan.

Methods

Study design, setting, and ethical considerations

Since 1969, the Japan Liver Cancer Association (JLCA) has been performing nationwide surveys of patients with primary liver cancer. Patients are registered and followed up as previously reported.¹⁶ The number of registered institutions that took part in the survey in years 2014-2015 was 516. A total of 63,163 patients were surveyed in years 2014-2015 across these institutions, accounting for approximately one-third of all patients with HCC treated in Japan.¹⁷ We set our study period from 2008 to 2017 based on the availability of recent treatment data.

The Ethics Committee of the Japan Institute for Health Security approved the study protocol (approval number JIHS-S-004367-05). This study adhered to the Declaration of Helsinki. Informed consent was waived due to the retrospective nature of this study. Data from patients with HCC were collected and registered with the approval of each institution participating in the nationwide survey.

Study patients

The inclusion criteria were: a Child–Pugh classification of A or B, no extrahepatic metastasis, no vascular invasion, three or fewer tumors, maximum tumor diameter of ≤ 3 cm, and actual treatment of LR or RFA monotherapy.

Definitions

The Japan integrated stage score comprised the sum of points for the two variables of the Japanese TNM classification and Child–Pugh classification.¹⁸ The liver damage grade was calculated according to serum bilirubin (mg/dL), serum albumin (g/dL), prothrombin activity (%), indocyanine green retention rate at 15 min, and ascites, similar to the Child–Pugh classification.¹⁹ The albumin–bilirubin (ALBI) grade was classified using the linear predictor = $(\log_{10} \text{bilirubin } [\mu\text{mol/L}] \times 0.66) + (\text{albumin } [\text{g/L}] \times -0.085)$.²⁰

Overall survival (OS) is defined as the time from treatment initiation to death from any cause. Patients who are still alive or who are lost to follow-up will be censored at the time of the last survival confirmation date.

Statistical analysis

The Kaplan–Meier method was employed to estimate OS, and confidence intervals (CI) were calculated following the Greenwood formula. A log-rank test was conducted to compare the survival curves of the two groups. *P*-values of <0.05 indicated significance. To address potential confounding factors in comparing the survival curves of two groups, 1:1 propensity score matching (PSM) with the nearest neighbor method was conducted using logistic regression. The TNM stage by JLCA was considered a stratification factor, with age, sex, mALBI grade 1, Child–Pugh A, number of tumors, maximum tumor diameter, hepatitis C virus antibody (HCV-Ab) positivity, and hepatitis B virus antigen (HBsAg) positivity as covariates, for the estimation of propensity scores. The *glm* function from the *statsmodels* was

employed to conduct the PSM process with a caliper width of 0.1, ensuring closer matching between treatment and control units and improving covariate balance. The analysis excluded cases with missing data in any of the variables used for propensity score estimation. To evaluate matching quality, we compared the standardized mean differences (SMD) of covariates between the two groups after matching. An SMD that is <0.1 indicated a negligible difference in the mean or prevalence of a covariate between treatment groups, despite the absence of a universally agreed-upon criterion regarding the SMD threshold to be used to indicate an important imbalance.^{21,22} Python (version 3.9.7) with the libraries *lifelines*, *scikit-survival*, and *statsmodels* was used for statistical analyses.

DL modeling

Python (version 3.9.7) was used to perform machine learning models, implemented in DSM. The underlying model of DSM is implemented in *PyTorch*. Table 1 demonstrates the variables included in the AI model

RDSM models the underlying event's outcome distribution as a combination of some fixed parametric distributions. The parameters of the mixture of these distributions, as well as the mixing weights, are modeled using Neural Networks. Thus:

$$P(T = t|X = x) = \sum_k P(Z = t|X = x)P(T = t|X = x, Z = k). \quad (1)$$

Here, t indicates the time at which an event of interest took place, x denotes the set of

covariates; $P(T = t|X = x) = \text{softmax}(f(\Phi(x)))$ where f is a linear function and

$P(T = t|X = x, Z = k)$ is Weibull or Log-Normal with shape and scale parameterized as functions of the representation $\Phi(x)$.

$$\begin{aligned} \ln P(T = t|X = x) &= \ln \sum_k P(Z = t|X = x)P(T = t|X = x, Z = k) \\ &\geq \sum_k P(Z = t|X = x) \ln P(T = t|X = x, Z = k). \end{aligned} \quad (2)$$

In the original DSM model, inference was performed employing a lower bound to the maximum likelihood estimate (Equation [2]). RDSM extends this framework using recurrent neural networks (e.g., Long Short-Term Memory networks, Gated Recurrent Units) to learn dynamic representations, making it well-suited for survival analysis tasks with time-varying covariates.

Our hyperparameter setting includes 4 Log-Normal distributions as the underlying survival distributions, 100 hidden layers, 4 neurons in each hidden layer, RNN as the type of recurrent cell, and 0.00068 learning rate.

Preprocessing

Variables available preoperatively with <20% missing data were used as model inputs. Missing values for continuous variables were imputed using the median, followed by normalization with a MinMaxScaler. Categorical variables were one-hot encoded, with missing values considered as absence (i.e., encoded as 0).

Evaluation method

To assess model performance, we employed 10-fold cross-validation. The dataset was randomly partitioned into 10 equal subsets (folds), comprising 9 folds for training and 1 fold for validation in each iteration. This process was repeated 10 times, ensuring that each fold served as the validation set exactly once. To prevent data leakage, each validation fold was kept completely separate from the training data during its corresponding iteration. This ensured that model performance was assessed on unseen data in each fold, thereby maintaining validation process integrity.

In the validation data, the 5-year survival rates for LR and RFA were predicted based on a trained model for each patient, and the treatment with the higher predicted survival rate was defined as the recommended treatment. Further, patients were classified into recommended and nonrecommended treatment groups based on whether they received the recommended therapy.

Metrics

The algorithm performance was evaluated using the concordance index (C-index) and OS. The C-index, widely used in survival analysis,²³ measures the concordance between predicted risk scores and actual survival times. The C-index is the proportion of comparable pairs where the predicted risk ranking matched observed survival outcomes, ranging from 0.5 (random concordance) to 1.0 (perfect discrimination).

To assess the survival benefit of the recommended treatment, OS was compared between the recommended and nonrecommended treatment groups.

Results

Participant selection

A total of 161,558 patients with HCC were identified from the survey. Of them, 59,250 patients with extrahepatic metastasis or unknown status; 26,238 patients with Child–Pugh C or unknown liver function status; 20,461 patients with tumor measuring >3 cm in diameter or unknown status; 19,367 patients with unknown treatment or treatments other than RFA or LR, not for curative intent, combined treatments, or not percutaneous; 11,644 patients with vascular invasion or unknown status; and 5,640 patients with more than three tumors or unknown status were excluded. Finally, the analysis included 18,958 patients (Figure S1).

Participant characteristics

Male patients were more common in those receiving LR than those with RFA (72.0% vs. 60.5%). Patients with modified albumin–bilirubin (mALBI) grade 1 (63.6% vs. 41.1%) and those with Child–Pugh A were more frequent in patients receiving LR than those with RFA (92.8% vs. 81.3%). Patients receiving LR more frequently demonstrated early enhancement and washout in the computed tomography (CT) images (91.2% vs. 82.1% and 84.2 vs. 82.6%), indicating better tumor differentiation. Patients with a single tumor were more frequent in those receiving LR (85.5% vs. 81.3%). Concerning viral infection of the background liver, patients with positive HCV-Ab and positive HBsAg were more frequent in those receiving RFA (62.1% vs. 49.0%) and LR (18.6% vs. 11.2%), respectively (Table 2).

Survival outcomes based on the Kaplan–Meier method

The 5-year survival rate was longer in patients receiving LR than those with RFA (81.4% [95%CI: 80.3%, 82.5%] vs. 73.1% [95%CI: 71.7%, 74.5%]; $P < 0.005$, respectively) (Figure S2).

Model evaluation

In the validation data, the trained RDSM (hereafter “our model”) achieved a C-index of 0.68, indicating successful training and adequate predictive performance.

Treatment recommendation based on DL modeling

For the validation data, our model predicted 5-year survival rates for LR and RFA for each patient. Of 18,958 patients analyzed, 3,199 patients were recommended for RFA and 15,759 for LR. Furthermore, 10,071 patients received the recommended treatment, whereas 8,887 patients did not.

Table 3 shows the baseline characteristics of patients recommended to undergo RFA or LR. Male patients were slightly more frequent in patients recommended for LR than those recommended for RFA (67.6% vs. 64.2%). Patients with mALBI grade 1 and Child–Pugh A patients were more common in those recommended for RFA (94.7% vs. 45.5% and 99.2 vs. 85.4%, respectively). Patients recommended for LR more commonly showed early enhancement and washout in the CT images (90.4% vs. 71.5% and 85.2 vs. 75.1%, respectively), indicating better tumor differentiation. Patients with a single tumor were more

frequent in those recommended for RFA (91.1% vs. 82.2%). Concerning the viral infection of the background liver, patients with positive HCV-Ab and positive HBsAg were more frequent in those recommended for LR (56.3% vs. 47.0%) and RFA (34.0% vs. 11.6%), respectively.

Comparing patients who received recommended treatment to those who did not

Table 4 demonstrates baseline patient characteristics of the recommended or nonrecommended treatment groups. Male patients were more common in the recommended treatment than in the nonrecommended treatment groups (71.0% vs. 62.6%). Patients with mALBI grade 1 and Child–Pugh A were more common in the recommended treatment than in the nonrecommended treatment groups (60.9% vs. 45.8% and 92.4 vs. 82.5%, respectively). Patients in the recommended treatment group more frequently showed early enhancement and washout in the CT images (89.9% vs. 84.2% and 84.6 vs. 82.2%, respectively). Patients with a single tumor were more common in patients who underwent recommended treatment (84.8% vs. 82.4%). Patients with positive HCV-Ab were more frequent in the nonrecommended treatment group (58.2% vs. 51.6%).

Survival outcomes based on DL modeling before PSM

Patients in the recommended treatment group demonstrated significantly longer survival rates compared with those in the nonrecommended treatment group (5-year survival rate [95%CI] 81.2% [80.1%, 82.3%] vs. 73.9% [72.5%, 75.2%]; $P < 0.005$) (Figure 1A).

Survival outcomes based on DL modeling after PSM

We successfully matched 7,099 patients in the recommended treatment group (70.5%) to 7,099 patients in the nonrecommended treatment group (79.9%). The SMDs for all covariates were <0.1 , indicating a good balance (Table 5). The OS in the recommended treatment group was longer than that in the nonrecommended treatment group (5-year survival rate [95%CI] 81.9% [80.6%, 83.2%] vs. 76.7% [75.3%, 78.1%]; $P < 0.005$) (Figure 1B).

Subgroup analysis for patients undergoing RFA where LR was recommended based on treatment types

A total of 6,966 (84.5%) patients underwent RFA where LR was recommended. Patients who underwent LR where RFA was recommended commonly showed early enhancement and washout in CT images (85.9% vs. 61.7% and 84.1 vs. 74.3%, respectively) compared with patients who received RFA as recommended (Table 6).

Treatment recommendation based on DL modeling in Child–Pugh A patients

Of 16,638 Child–Pugh A patients, 3,174 patients were recommended for RFA and 13,464 for LR . Furthermore, 9,305 patients received the recommended treatment, whereas 7,333 patients did not.

Supplementary table 1 presents the baseline characteristics of Child–Pugh A patients recommended for RFA or LR. Male patients were slightly more frequent among patients recommended for LR than among those recommended for RFA (68.9% vs. 64.1%). Patients with modified albumin–bilirubin (mALBI) grade 1 were more common among those recommended for RFA (95.0% vs. 51.6%). Patients recommended for LR more commonly demonstrated early enhancement and washout in the CT images (91.1% vs. 71.6% and 85.7 vs. 75.2%, respectively), indicating better tumor differentiation. Patients with a single tumor were more frequent among those recommended for RFA (91.3% vs. 82.4%). Regarding the viral infection of the background liver, HCV-Ab-positive and HBsAg-positive patients were more frequent among those recommended for LR (55.9% vs. 46.9%) and RFA (33.8% vs. 12.0%), respectively.

Comparison of patients who received the recommended treatment with those who did not received it among Child–Pugh A patients

Supplementary table 2 shows the baseline patient characteristics of the recommended or nonrecommended treatment groups among Child–Pugh A patients. Male patients were more common in the recommended treatment than in the nonrecommended treatment groups (71.2% vs. 63.8%). Patients with mALBI grade 1 were more common in the recommended treatment than in the nonrecommended treatment groups (64.1% vs. 54.5%). Patients in the recommended treatment group more frequently demonstrated early enhancement and washout in the CT images (89.9% vs. 84.2% and 84.9 vs. 82.2%, respectively). Patients with a single tumor were more common in the recommended treatment group (84.8% vs. 83.1%). HCV-Ab-positive patients were more frequent in the nonrecommended treatment group (57.5% vs. 51.6%).

Survival outcomes based on DL modeling before PSM among Child–Pugh A patients

Patients in the recommended treatment group exhibited significantly longer survival rates than those in the nonrecommended treatment group (5-year survival rate [95% confidence interval (CI)] 82.2% [81.0%, 83.3%] vs. 78.1% [76.7%, 79.5%]; $P < 0.005$) (Figure 2A).

Survival outcomes based on DL modeling after PSM among Child–Pugh A patients

We successfully matched 4,660 patients in the recommended treatment group (50.1%) to 4,660 patients in the nonrecommended treatment group (63.5%). The SMDs for all covariates were <0.1 , indicating good balance (Supplementary table 3). The OS in the recommended treatment group was longer than

that in the nonrecommended treatment group (5-year survival rate [95% CI] 82.0% [80.3%, 83.5%] vs. 79.1% [77.3%, 80.8%]; $P < 0.005$) (Figure 2B).

Discussion

To the best of our knowledge, our study is the first to report that the DL model accurately recommended treatments for patients with up to three HCCs measuring ≤ 3 cm. Patients in the recommended treatment group demonstrated significantly longer survival than those in the nonrecommended treatment group, even after PSM. Several recent reports have highlighted the use of DL models in HCC treatments; however, they are mainly for diagnosis and predicting prognosis.⁹

LR and RFA are curative treatment modalities for patients with HCC with up to three HCCs measuring ≤ 3 cm. However, considering the fundamental differences between the two treatment modalities, individual patients may benefit more from one approach than the other, according to tumor characteristics and liver function reserve. The application of the present DL model would benefit clinicians in allocating appropriate treatments for patients. In our DL model, LR was recommended for 84.5% of the patients who underwent RFA, and these patients more often showed early enhancement and washout in CT images compared with those recommended for RFA. The OS difference among patients in the recommended treatment group indicates that applying this DL model may improve survival in some patients undergoing RFA in daily clinical practice. This difference may relate to CT enhancement patterns. Although future prospective validation is warranted, the model's application appears promising.

We further validated the utility of the DL model in the Child–Pugh A population. Results showed that patients in the recommended treatment group had significantly better survival than those in the nonrecommended treatment group. Moreover, ALBI grade and CT imaging findings were significantly different between these groups. It has been initially reported that ALBI grade can classify Child–Pugh A patients, and this fact appeared important in the DL model. CT enhancement patterns also appeared important in the DL model for deciding treatment allocation. Tumor enhancement pattern has been reported to be associated with tumor differentiation.²⁴ Nevertheless, tumor differentiation was not included in the variables of the model, which needs to be investigated in future studies.

This study has some limitations. Our retrospective study is subject to selection bias, which we attempt to minimize using PSM, although other biases may remain unaccounted for. Our DL model's application should be validated in future prospective trials. Our DL model did not incorporate the tumor location. Administering RFA for peribiliary HCC is difficult^{25,26}; therefore, future DL models should incorporate the tumor location. Further, we did not consider postprocedural complication rates, which are crucial in clinical decision making, nor include patients at an advanced stage of disease. Future DL models should include patients with more advanced HCC eligible for chemotherapy and transarterial chemoembolization to truly allocate treatments for all patients with HCC.

In conclusion, this study is the first report that the DL model effectively recommended treatments accurately for patients with up to three HCCs measuring ≤ 3 cm. Patients in the recommended treatment group demonstrated significantly longer survival than those in the nonrecommended treatment group.

Declarations

Funding

Small Business Innovation Research (SBIR).

Author Contribution

TK. and YY. and NK wrote the main manuscript text and TK, AK, YK prepared figures. All authors reviewed the manuscript

Data Availability

The data that support the findings of this study are available from the Japan Liver Cancer Association but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of the Japan Liver Cancer Association.

References

1. Tan EY, Danpanichkul P, Yong JN, et al. Liver cancer in 2021: Global Burden of Disease study. *J Hepatol.* 2025;82(5):851-860.
2. Hasegawa K, Takemura N, Yamashita T, et al. Clinical Practice Guidelines for Hepatocellular Carcinoma: The Japan Society of Hepatology 2021 version (5th JSH-HCC Guidelines). *Hepatol Res* 2023;53 (5):383-390.
3. European Association for the Study of the Liver. EASL Clinical Practice Guidelines on the management of hepatocellular carcinoma. *J Hepatol* 2025;82(2):315-374.
4. 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.
5. Chen MS, Li JQ, Zheng Y, et al. A prospective randomized trial comparing percutaneous local ablative therapy and partial hepatectomy for small hepatocellular carcinoma. *Ann Surg* 2006;243(3):321–328.
6. Huang J, Yan L, Cheng Z, et al. A randomized trial comparing radiofrequency ablation and surgical resection for HCC conforming to the Milan criteria. *Annals of surgery* 2010;252(6):903–912.

7. Feng K, Yan J, Li X, et al. A randomized controlled trial of radiofrequency ablation and surgical resection in the treatment of small hepatocellular carcinoma. *J Hepatol* 2012;57(4):794–802.
8. Takayama T, Hasegawa K, Izumi N, et al. Surgery versus Radiofrequency Ablation for Small Hepatocellular Carcinoma: A Randomized Controlled Trial (SURF Trial). *Liver Cancer* 2021;11(3):209-218.
9. Chatzipanagiotou OP, Loukas C, Vailas M, et al. Artificial intelligence in hepatocellular carcinoma diagnosis: a comprehensive review of current literature. *J Gastroenterol Hepatol* 2024;39(10):1994-2005.
10. Han JW, Lee SK, Kwon JH, et al. A Machine Learning Algorithm Facilitates Prognosis Prediction and Treatment Selection for Barcelona Clinic Liver Cancer Stage C Hepatocellular Carcinoma. *Clin Cancer Res* 2024;30(13):2812-2821.
11. Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Med Res Methodol* 2018;18(1):24.
12. Nagpal C, Li X, Dubrawski A. Deep Survival Machines: Fully Parametric Survival Regression and Representation Learning for Censored Data With Competing Risks. *IEEE J Biomed Health Inform* 2021;25(8):3163-3175.
13. Nagpal C, Vincent J, Dubrawski A. Deep Parametric Time-to-Event Regression with Time-Varying Covariates. *Proceedings of AAAI Spring Symposium on Survival Prediction - Algorithms, Challenges, and Applications* 2021. PMLR 2021;146: 184-193.
14. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput* 1997;9(8):1735-1780.
15. Cho K, Merriënboer BV, Gulcehre C, et al. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv* 2014; 1406.1078.
16. Hasegawa K, Kokudo N, Makuuchi M, et al. Comparison of resection and ablation for hepatocellular carcinoma: a cohort study based on a Japanese nationwide survey. *J Hepatol* 2013;58(4):724-729.
17. Iijima H, Kudo M, Kubo S, et al. Report of the 23rd nationwide follow-up survey of primary liver cancer in Japan (2014-2015). *Hepatol Res* 2023;53(10):895-959.
18. Luo KZ, Itamoto T, Amano H, et al. Comparative study of the Japan Integrated Stage (JIS) and modified JIS score as a predictor of survival after hepatectomy for hepatocellular carcinoma. *J Gastroenterol* 2008;43(5):369-377.
19. Omagari K, Ohba K, Kadokawa Y, et al. Comparison of the grade evaluated by “Liver damage” of Liver Cancer Study Group of Japan and Child-Pugh classification in patients with hepatocellular carcinoma. *Hepatol Res* 2006;34(4):266-272.
20. Johnson PJ, Berhane S, Kagebayashi C, et al. Assessment of liver function in patients with hepatocellular carcinoma: a new evidence-based approach-the ALBI grade. *J Clin Oncol* 2015;33(6):550-558.
21. Austin, PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res* 2011;46(3):399-424.

22. Romeo M, Dallio M, Napolitano C, et al. Clinical Applications of Artificial Intelligence (AI) in Human Cancer: Is It Time to Update the Diagnostic and Predictive Models in Managing Hepatocellular Carcinoma (HCC)? *Diagnostics (Basel)* 2025;15:252.

23. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Stat Med* 2011;30(10):1105-1117.

24. Yoon SH, Lee JM, So YH, et al. Multiphasic MDCT enhancement pattern of hepatocellular carcinoma smaller than 3 cm in diameter: tumor size and cellular differentiation. *AJR Am J Roentgenol.* 2009;193(6):W482-489.

25. Cucchetti A, Elshaarawy O, Han G, et al. 'Potentially curative therapies' for hepatocellular carcinoma: how many patients can actually be cured? *Br J Cancer* 2023;128(9):1665-1671.

26. Cui J, Sui X, Liu K, et al. Radiofrequency ablation for peribiliary hepatocellular carcinoma: propensity score matching analysis. *Insights Imaging* 2025;16(1):45.

Tables

Table 1. Included variables in the artificial intelligence model

Treatment	Sex
Age	Stage based on image findings
Alpha-fetoprotein (ng/ml)	Maximum tumor diameter based on
Japan's integrated stage score	image findings
Child–Pugh grade	Bilobar or unilobar for multiple tumors
Hepatitis B surface antigen	Tumor number based on image findings
Hepatitis C virus antibody	A combination of hepatitis B surface
Protein induced by vitamin K absence or	antigen and hepatitis C virus antibody
antagonist-II (mAU/ml)	positivity
Modified albumin–bilirubin grade	Liver damage grade
Modified albumin–bilirubin Score	Platelet count ($10^4/\text{mm}^3$)
History of heavy use of alcohol	Serum albumin level (g/dl)
Prothrombin activity	Serum bilirubin (mg/dl)
Wash out in the portal phase in contrast-	
enhanced computed tomography	
Early enhancement in the arterial phase	
of the contrast-enhanced computed	
tomography	

Table 2. Baseline characteristics according to treatment type

Treatment	RFA performed (n = 8,244)	LR performed (n = 10,714)	Difference* (95%CI)
Age	70.4 (9.6)	68.2 (9.6)	2.2 [-7.4, 11.8]
Sex (Male/Female)	4,990/3,254 (60.5/39.5)	7,718/2,996 (72.0/28.0)	-11.5 [-12.9, -10.2] /11.5 [10.2, 12.9]
mALBI grade 1	3,387 (41.1)	6,813 (63.6)	-22.5 [-23.9, -21.1]
Child-Pugh A	6,699 (81.3)	9,939 (92.8)	-11.5 [-12.5, -10.5]
Single tumor	6,699 (81.3)	9,164 (85.5)	-4.3 [-5.3, -3.2]
Maximum tumor diameter(cm)	1.7 (0.5)	2.1 (0.6)	-0.4 [-1.0, 0.2]
HCV-Ab positive	5,116 (62.1)	5,255 (49.0)	13.0 [11.6, 14.4]
HBsAg positive	927 (11.2)	1,994 (18.6)	-7.4 [-8.4, -6.4]
AFP (ng/mL)	15.7 (4.6)	16.5 (6.5)	1.0 [0.2, 5.5]
Platelet count (10 ⁴ /mm ³)	10.9 (1.7)	13.7 (1.6)	0.8 [0.5, 1.3]
Serum albumin (g/dL)	3.8 (0.5)	4.0 (0.5)	-0.3 [-0.8, 0.2]
Serum bilirubin (mg/dL)	0.8 (1.6)	0.7 (1.6)	1.1 [0.7, 1.8]
Prothrombin activity (%)	83.6 (16.1)	88.0 (17.7)	-4.4 [-21.5, 12.7]
Early enhancement**	6,772 (82.1)	9,766 (91.2)	-9.0 [-10.0, -8.0]
Washout ***	6,806 (82.6)	9,025 (84.2)	-1.7 [-2.7, -0.6]
Data other than AFP, platelet count, and serum bilirubin are reported as mean values with standard deviations or as numbers (%). AFP, platelet count, and serum bilirubin are reported as geometric mean with geometric standard deviation. CI, confidence interval; LR, liver resection; RFA, radiofrequency ablation; mALBI, modified albumin–bilirubin; HCV-Ab, hepatitis C virus antibody; HBsAg, hepatitis B virus antigen; AFP, alpha-fetoprotein. * The geometric mean ratios were used for AFP, platelet count, and serum bilirubin, whereas mean differences were used for all other variables, ** arterial phase in contrast-enhanced computed tomography, *** portal phase in contrast-enhanced computed tomography			

Table 3. Baseline characteristics according to recommended treatment type

Treatment	RFA recommended (n = 3,199)	LR recommended (n = 15,759)	Difference* (95%CI)
Age	66.1 (10.6)	69.8 (9.3)	-3.7 [-16.4, 9.0]
Sex (Male/Female)	2,055/1,144 (64.2/35.8)	10,653/51,06 (67.6/32.4)	-3.4 [-5.2, -1.5] /3.4 [1.5, 5.2]
mALBI grade 1	3,031 (94.7)	7,169 (45.5)	49.3 [48.2, 50.4]
Child–Pugh A	3,174 (99.2)	13,464 (85.4)	13.8 [13.2, 14.4]
Single tumor	2,914 (91.1)	12,949 (82.2)	8.9 [7.8, 10.1]
Maximum tumor diameter(cm)	1.8 (0.6)	1.9 (0.6)	-0.2 [-1.0, 0.7]
HCV-Ab positive	1,504 (47.0)	8,867 (56.3)	-9.3 [-11.1, -7.4]
HBsAg positive	1,088 (34.0)	1,833 (11.6)	22.4 [20.7, 24.1]
AFP (ng/mL)	14.3 (6.2)	16.6 (5.6)	0.9 [0.1, 8.5]
Platelet count (x10 ⁴ /mm ³)	14.1 (1.6)	12.1 (1.7)	1.2 [0.6, 2.3]
Serum albumin (g/dL)	4.3 (0.3)	3.8 (0.5)	0.4 [-0.2, 1.1]
Serum bilirubin (mg/dL)	0.7 (1.5)	0.8 (1.6)	0.9 [0.5, 1.6]
Prothrombin activity (%)	92.5 (13.8)	84.8 (17.5)	7.7 [-14.8, 30.2]
Early enhancement**	2,286 (71.5)	14,252 (90.4)	-19.0 [-20.6, -17.3]
Washout ***	2,403 (75.1)	13,428 (85.2)	-10.1 [-11.7, -8.5]
Data other than AFP, platelet count, and serum bilirubin are reported as mean values with standard deviations or as numbers (%). AFP, platelet count, and serum bilirubin are reported as geometric mean with geometric standard deviation. LR, liver resection; RFA, radiofrequency ablation; mALBI, modified albumin–bilirubin; HCV-Ab, hepatitis C virus antibody; HBsAg, hepatitis B virus antigen; AFP, alpha-fetoprotein. * The geometric mean ratios were used for AFP, platelet count, and serum bilirubin, whereas mean differences were used for all other variables,** arterial phase in contrast-enhanced computed tomography, *** portal phase in contrast-enhanced computed tomography			

Table 4. Baseline characteristics according to recommended treatment or not recommended treatment

Treatment	Recommended treatment (n =10,071)	Not recommended treatment (n = 8,887)	Difference* (95%CI)
Age	69.0 (9.4)	69.4 (9.9)	-0.4 [-10.0, 9.2]
Sex (Male/Female)	7,147/2,924 (71.0/29.0)	5,561/3,326 (62.6/37.4)	8.4 [7.1, 9.7] /-8.4 [-9.7, -7.1]
mALBI grade 1	6,130 (60.9)	4,070 (45.8)	15.1 [13.7, 16.5]
Child–Pugh A	9,305 (92.4)	7,333 (82.5)	9.9 [8.9, 10.8]
Single tumor	8,538 (84.8)	7,325 (82.4)	2.4 [1.3, 3.4]
Maximum tumor diameter(cm)	2.0 (0.6)	1.8 (0.6)	0.3 [-0.3, 0.9]
HCV-Ab positive	5,197 (51.6)	5,174 (58.2)	-6.6 [-8.0, -5.2]
HBsAg positive	1,562 (15.5)	1,359 (15.3)	0.2 [-0.8, 1.2]
AFP (ng/mL)	15.6 (6.1)	16.8 (5.2)	0.9 [0.2, 5.3]
Platelet count (10 ⁴ /mm ³)	13.4 (1.6)	11.3 (1.7)	1.2 [0.7, 2.0]
Serum albumin (g/dL)	4.0 (0.5)	3.8 (0.5)	0.2 [-0.3, 0.7]
Serum bilirubin (mg/dL)	0.7 (1.6)	0.8 (1.6)	0.9 [0.6, 1.4]
Prothrombin activity (%)	87.3 (17.6)	84.7 (16.6)	2.6 [-14.4, 19.7]
Early enhancement**	9,058 (89.9)	7,480 (84.2)	5.8 [4.8, 6.7]
Washout ***	8,522 (84.6)	7,309 (82.2)	2.4 [1.3, 3.4]
Data other than AFP, platelet count, and serum bilirubin are reported as mean values with standard deviations or as numbers (%). AFP, platelet count, and serum bilirubin are reported as geometric mean with geometric standard deviation. LR, liver resection; RFA, radiofrequency ablation; mALBI, modified albumin–bilirubin; HCV-Ab, hepatitis C virus antibody; HBsAg, hepatitis B virus antigen. * The geometric mean ratios were used for AFP, platelet count, and serum bilirubin, whereas mean differences were used for all other variables, ** arterial phase in contrast-enhanced computed tomography, *** portal phase in contrast-enhanced computed tomography.			

Table 5. Baseline characteristics of the 14,178 patients with hepatocellular carcinoma matched by propensity score according to recommended treatment or not recommended treatment

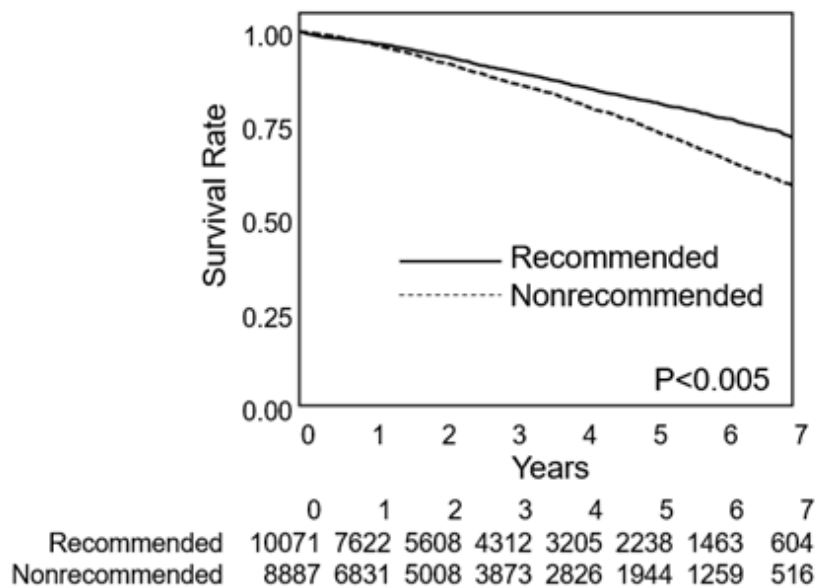
Treatment	Recommended treatment (n = 7,099)	Not recommended treatment (n = 7,099)	Mean difference (95%CI)	SMD
Age	69.3 (9.0)	68.3 (9.9)	0.9 [-8.5, 10.3]	0.099
Sex (Male/Female)	4,773/2,326 (67.2/32.8)	4,713/2,386 (66.4/33.6)	0.8 [-0.7, 2.4] /-0.8 -2.4, 0.7]	0.018/-0.018
mALBI grade 1	4,015 (56.6)	3,845 (54.2)	2.4 [0.8, 4.0]	0.048
Child–Pugh A	6,419 (90.4)	6,334 (89.2)	1.2 [0.2, 2.2]	0.04
Number of tumors	1.2 (0.5)	1.2 (0.5)	0.0 [-0.5, 0.5]	0.008
Maximum tumor diameter(cm)	1.8 (0.6)	1.8 (0.6)	0.0 [-0.6, 0.6]	0.008
HCV-Ab positive	3,836 (54.0)	3,950 (55.6)	-1.6 [-3.2, 0.0]	-0.032
HBsAg positive	1,236 (17.4)	1,190 (16.8)	0.6 [-0.6, 1.9]	0.017
Data are reported as mean values with standard deviations or as numbers (%). LR, liver resection; RFA, radiofrequency ablation; mALBI, modified albumin–bilirubin; HCV-Ab, hepatitis C virus antibody; HBsAg, hepatitis B virus antigen.				

Table 6. Baseline characteristics of patients who underwent RFA according to the recommended treatment based on the DL algorithm

Treatment	RFA performed RFA recommended (n = 1,278)	RFA performed LR recommended (n = 6,966)	Difference* (95%CI)
Age	68.5 (10.5)	70.7 (9.4)	-2.2 [-15.4, 10.9]
Sex (Male/Female)	742/536 (58.1/41.9)	4,248/2,718 (61.0/39.0)	-2.9 [-5.9, 0.0]/2.9 [-0.0, 5.9]
ALBI grade 1	1,174 (91.9)	2,213 (31.8)	60.1 [58.2, 61.9]
Child–Pugh A	1,270 (99.4)	5,429 (77.9)	21.4 [20.4, 22.5]
Single tumor	1,144 (89.5)	5,555 (79.7)	9.8 [7.8, 11.7]
Maximum tumor diameter(cm)	1.5 (0.5)	1.7 (0.5)	-0.2 [-0.9, 0.6]
HCV-Ab positive	723 (56.6)	4,393 (63.1)	-6.5 [-9.4, -3.5]
HBsAg positive	328 (25.7)	599 (8.6)	17.1 [14.6, 19.5]
AFP (ng/mL)	11.2 (4.4)	16.8 (4.6)	0.7 [0.1, 5.3]
Platelet count (10 ⁴ /mm ³)	13.0 (1.6)	10.5 (1.7)	1.2 [0.6, 2.5]
Serum albumin (g/dL)	4.2 (0.3)	3.7 (0.5)	0.5 [-0.2, 1.2]
Serum bilirubin (mg/dL)	0.7 (1.5)	0.9 (1.6)	0.8 [0.4, 1.5]
Prothrombin activity (%)	91.0 (12.6)	82.3 (16.3)	8.7 [-13.0, 30.5]
Early enhancement**	789 (61.7)	5,983 (85.9)	-24.2 [-26.9, -21.4]
Washout ***	950 (74.3)	5,856 (84.1)	-9.7 [-12.3, -7.2]
Data other than AFP, platelet count, and serum bilirubin are reported as mean values with standard deviations or as numbers (%). AFP, platelet count, and serum bilirubin are reported as geometric mean with geometric standard deviation. LR, liver resection; RFA, radiofrequency ablation; mALBI, modified albumin–bilirubin; HCV-Ab, hepatitis C virus antibody; HBsAg, hepatitis B virus antigen; AFP, alpha-fetoprotein. * The geometric mean ratios were used for AFP, platelet count, and serum bilirubin, whereas mean differences were used for all other variables, ** arterial phase in contrast-enhanced computed tomography, *** portal phase in contrast-enhanced computed tomography			

Figures

(A)



(B)

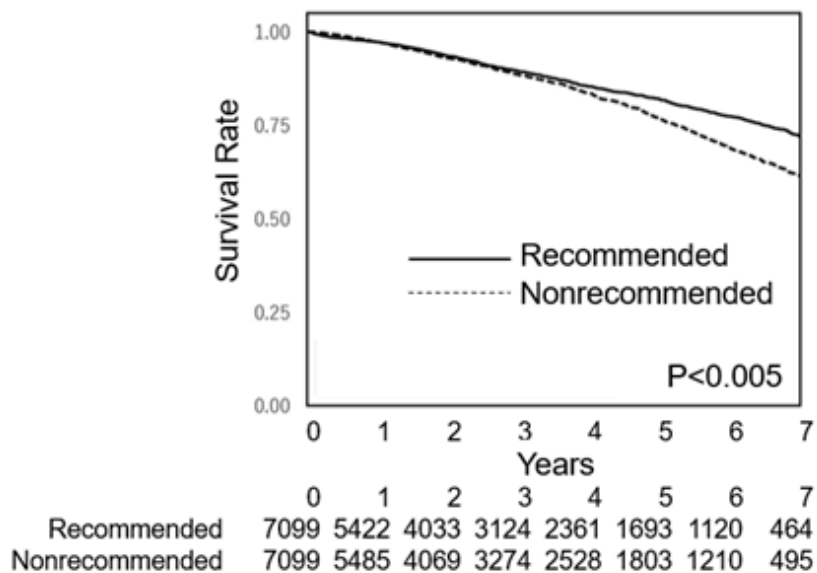
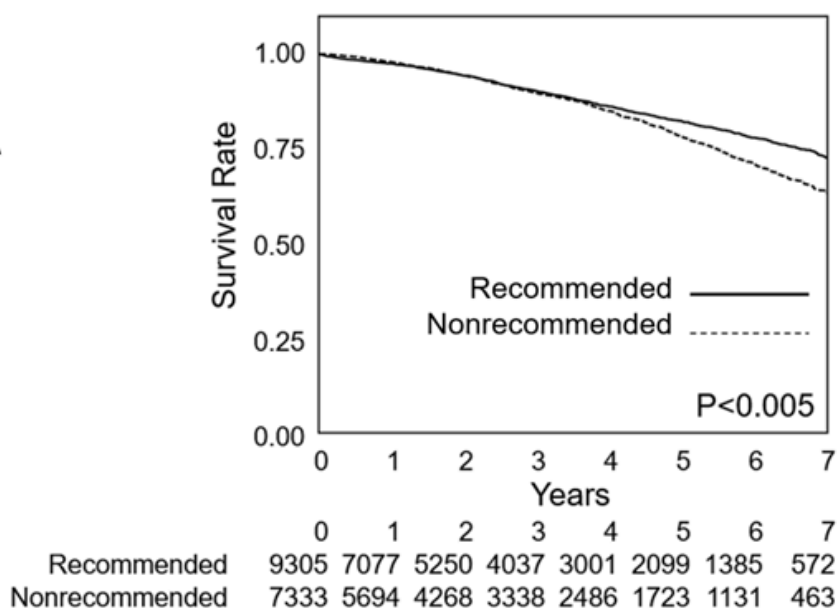


Figure 1

(A) Overall survival of patients in the recommended and nonrecommended treatment groups. (B) after propensity score matching.

A



B

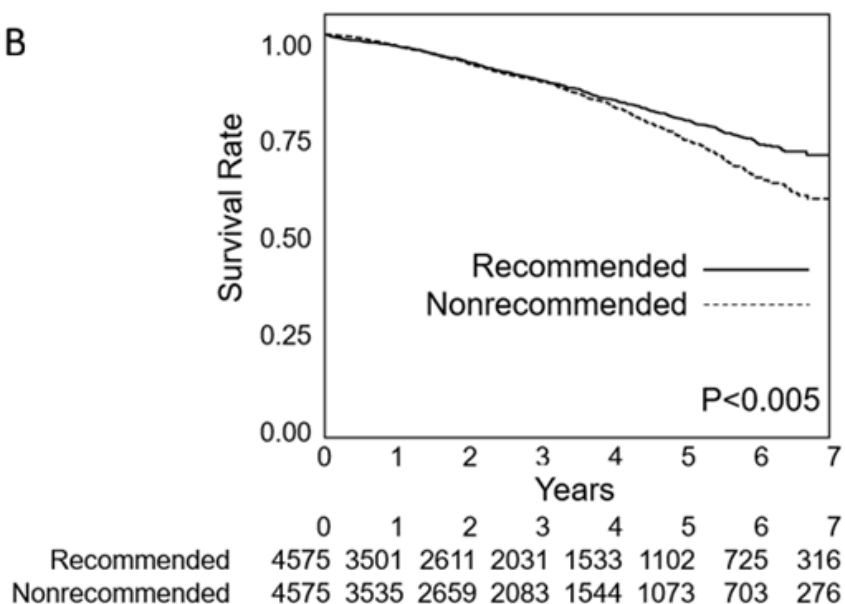


Figure 2

A) Overall survival of patients in the recommended and nonrecommended treatment groups among Child-Pugh A patients.

Supplementary Files

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

- [Supplementary.docx](#)