

Invited Editorial**Unlocking the future: machine learning sheds light on prognostication for early-stage hepatocellular carcinoma**

Junlong Dai^{1,2}, Jimmy Che-To Lai^{1,2}, Grace Lai-Hung Wong^{1,2}, Terry Cheuk-Fung Yip^{1,2}

¹Medical Data Analytics Center, Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong

²State Key Laboratory of Digestive Disease, Institute of Digestive Disease, The Chinese University of Hong Kong, Hong Kong

Running title: Machine Learning for Prognosis of Early HCC

Correspondence:

Dr Terry Cheuk-Fung Yip, PhD.

Department of Medicine and Therapeutics, 9/F, Lui Che Woo Clinical Sciences Building, Prince of Wales Hospital, 30-32 Ngan Shing Street, Shatin, Hong Kong

Tel: +852 3505 3125 Fax: +852 2637 3852

Email: tcfyip@cuhk.edu.hk

Total word count: 1,415

No. of figures: 1

Financial support: None declared.

List of abbreviations: ALBI=albumin-bilirubin, AST=aspartate aminotransferase, AUC=area under the receiver operating characteristic curve, BCLC=Barcelona Clinic of Liver Cancer, FIB-4=fibrosis 4, HCC=hepatocellular carcinoma, LASSO=Least Absolute Shrinkage and Selection Operator, PNI=Prognostic Nutritional Index.

Conflict of interest statement:

Junlon Dai declares that he has no conflicts of interest.

Jimmy Lai has served as a speaker for Gilead Sciences and an advisory board committee for Gilead Sciences and Boehringer Ingelheim.

Grace Wong has served as an advisory committee member for AstraZeneca, Gilead Sciences and Janssen, and as a speaker for Abbott, AbbVie, Ascleptis, Bristol-Myers Squibb, Echosens, Gilead Sciences, Janssen, and Roche. She has also received a research grant from Gilead Sciences.

Terry Yip has served as an advisory committee member and a speaker for Gilead Sciences.

Patients with hepatocellular carcinoma (HCC) have different prognoses with regard to the staging of the disease. The Barcelona Clinic of Liver Cancer (BCLC) algorithm is the most widely used staging system categorizing patients with HCC into five different stages. By considering the tumor burden, liver function and performance status of the patient, the BCLC staging guides respective treatment modality across the stages and also predicts the prognosis of HCC. Among the five stages, early-stage HCC (*i.e.* BCLC stage 0-A) has the best prognosis.^{1,2} Apart from the BCLC staging, there are additional factors that predict the prognosis of early-stage HCC. For instance, a higher level of alpha-fetoprotein predicts HCC recurrence and poorer survival. The presence of satellite lesions and microvascular invasion on histological assessment also prognosticates the risk of HCC recurrence. Treatment modality also implicates the prognosis with local ablative therapy resulting in a higher recurrence rate than surgical resection for larger (*e.g.* >3 cm) tumors or tumors located at specific sites in the liver.^{1,2} Besides, clinically significant portal hypertension predicts a higher risk of post-operative complications and lower long-term survival.³ Last but not least, a higher Child-Pugh score leads to poorer prognosis.^{1,2}

To better estimate the prognosis of early-stage HCC, the albumin-bilirubin (ALBI) score was developed.⁴ Various studies showed that the ALBI score, using serum albumin and bilirubin levels, objectively correlated with the degree of liver impairment and stratified patients across different Child-Pugh score and BCLC stages, and thus was validated in the prediction of survival, duration to recurrence and treatment tolerability.⁵ In the era of precision medicine, better risk prediction for patients with early-stage HCC is imperative to guide individualized management. Machine learning is a powerful tool which usually outperforms traditional statistical approaches in accurate risk prediction modelling,⁶ but its use in the prediction of prognosis of early-stage HCC remains to be determined.

In this Issue of *Clinical and Molecular Hepatology*, Ho *et al.* built and validated a risk score for predicting overall survival in patients with early-stage HCC using a single-center retrospective cohort of 1411 patients with BCLC stage 0-A HCC.⁷ More than 90% of them received curative treatment including liver transplantation. The cohort was split into training and validation cohorts in a ratio of 70:30. A total of 17 variables that reflected tumor burden and liver function reserve formed the basis for the regression model. Concerning model development, Ho *et al.* adopted two tactics: stepwise Cox regression (conventional method) and Least Absolute Shrinkage and Selection Operator (LASSO) Cox regression (machine learning [ML]-based method).

Generally, when searching for variables that significantly affect patient prognosis, univariate Cox analysis is performed to screen out less relevant variables, and a multivariable model is constructed to further confirm the independent association between variables and survival. However, this approach does not take into account the impact of multicollinearity between variables. Ignoring multicollinearity can lead to contradictory hazard ratios in univariate and multivariable Cox regression due to distorted models caused by the interdependence of variables. LASSO algorithm uses the L1 norm, *i.e.*, the sum of the absolute value of the regression coefficients, for shrinkage penalty. It reduces the magnitude of variable coefficients that have little prognostic power to the dependent variable. As a mathematical property of the L1 norm, the coefficients of some less important variables can be reduced to 0

while keeping the coefficients of important variables greater than 0 to reduce the number of covariates in the Cox regression, ultimately addressing multicollinearity while performing variable selection. Moreover, when there are many variables, the traditional discrete variable selection methods such as forward, backward, and stepwise selection methods are not efficient. Therefore, when multicollinearity between variables is suspected or the number of variables is greater than the sample size (e.g. screening genes that affect prognosis), LASSO Cox regression needs to be applied to select prognostic variables.

Ho *et al.* compared the performance of stepwise-based (CATS-IF) and LASSO-based risk scores (CATS-INF). The CATS-IF score involves 8 variables while the CATS-INF score involves 10. Parameters in common included: serum creatinine, age, alpha-fetoprotein, ALBI grade, curative treatment of HCC, single large HCC, lymphocyte-monocyte ratio, and the fibrosis-4 (FIB-4) index, which are repetitively proven to have a strong impact on survival.^{8,9} The difference is that the ML-based CATS-INF score added two new variables: one reflecting overall nutritious status called Prognostic Nutritional Index (PNI) and the other one is aspartate aminotransferase (AST). PNI is determined by assessing the nutritional status through albumin level and evaluating the immune status based on lymphocyte count. Recent research has found a series of indicators like the presence of sarcopenia and the abnormality of skeletal muscle index that can worsen liver cancer patients' prognosis.^{10,11} PNI has also been compared to other inflammatory indices in predicting HCC prognosis.¹² The authors concluded that the ML-based risk score achieved a slightly higher area under the receiver operating characteristic curve (AUC) than the conventional one (0.707 vs. 0.695), both outperformed other common prognostic scores including ALBI, FIB-4, AST to platelet ratio index, and MELD score.

Ho *et al.*'s study serves as a valuable reference for future research. Firstly, although the two developed risk scores exhibited numerically higher AUC compared to the ALBI score, it is important to also examine the 95% confidence interval of AUC and determine whether the difference in AUC between the scores is statistically significant. Additionally, it is also crucial to consider the clinical significance, taking into account the ease of calculating the ALBI score. Secondly, it is worth noting that although LASSO regression is theoretically capable of mitigating overfitting in the risk model by shrinking beta coefficients, a closer examination of the beta coefficients for CATS-INF and CATS-IF reveals that both risk scores assigned equal weights to each variable, except CATS-INF incorporating two additional variables of PNI and AST. This observation could potentially account for the marginally higher yet comparable AUC between CATS-INF and CATS-IF. In future studies, it may be valuable to explore the number of variables when the model can significantly benefit from LASSO variable selection. The use of other popular machine learning algorithms such as survival support vector machine, random survival forest, and extreme gradient boosting may also be explored.^{6,8}

Thirdly, among the patients with BCLC stage 0-A stage HCC in Ho *et al.*'s study, approximately 15% and 10% had single HCC >5 cm and received no therapy of curative intent, respectively. These two groups of patients showed worse survival and were thus two strong risk factors in the two risk scores. Although the current study had fewer patients from these groups, it would be beneficial for future studies to assess the performance of risk scores in these patient subgroups and determine the necessity of developing a separate risk score tailored to their risk profile. Fourthly, the training and validation cohorts for CATS-INF and CATS-IF came from the same medical center. It may be of interest to explore the

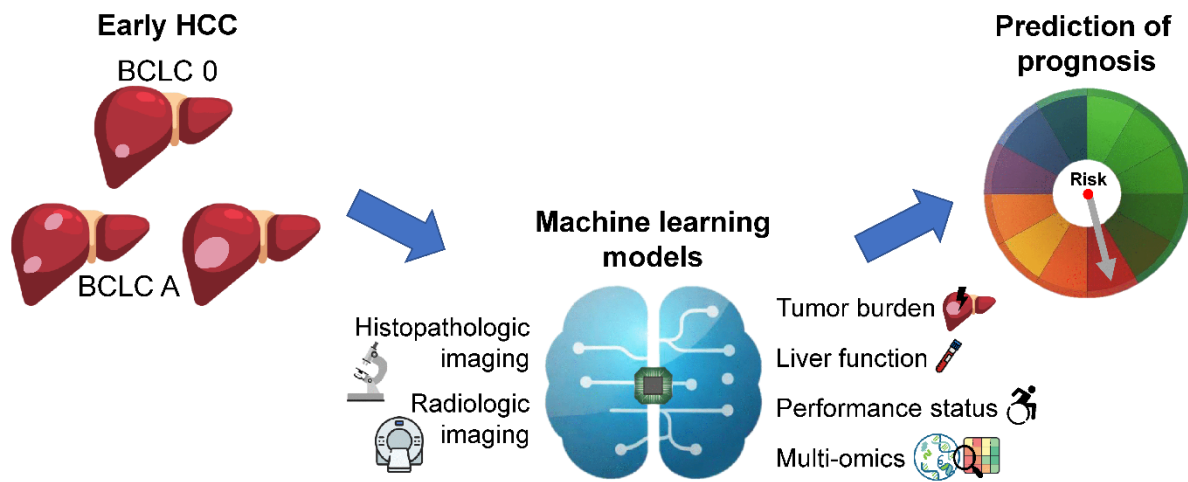
performance of the risk scores in an independent external validation cohort to gain insights into the generalizability and transportability of the risk scores. Finally, regarding the dichotomization of variables in the risk scores, exploring the difference in dichotomization and treating variables as continuous could be of interest to researchers aiming to assess their impact on model performance.¹³ Furthermore, since CATS-INF and CATS-IF incorporated certain index variables such as FIB-4, ALBI, and PNI, it would be worthwhile to examine and compare the performance of the risk scores when including these indices as a whole or when incorporating the individual components within each index.

We are now moving towards personalised prognostication for early-stage HCC with data-driven approaches (Figure 1). Multimodal prediction models leveraging comprehensive clinical, laboratory and radiomic parameters enhance prognostic accuracy and provide a more comprehensive assessment by considering not only tumour factors but also patient factors in terms of time series, which capture the evolving clinical conditions over time.^{14,15} Multimodal radiomics with image fusion and co-registration techniques integrate imaging modalities such as computed tomography, magnetic resonance imaging, positron emission tomography, and even ultrasonography to obtain complementary information about tumour characteristics before and after HCC treatment.¹⁶ All these advances facilitate dynamic prediction of HCC recurrence and modify subsequent risk of death. Future research remains adaptive to changing aetiologies of HCC, with decreasing incidences of HCC secondary to chronic viral hepatitis in the presence of potential antiviral treatments, and increasing incidences of HCC related to metabolic dysfunction-associated steatotic liver disease.¹⁷ Any additional benefits of the integration of novel biomarkers and liquid biopsies in further enhancing prognostic accuracy or enabling early detection of HCC recurrence are yet to be substantiated.

References

1. Reig M, Forner A, Rimola J, Ferrer-Fabrega J, Burrel M, Garcia-Criado A, et al. BCLC strategy for prognosis prediction and treatment recommendation: The 2022 update. *J Hepatol* 2022;76:681-693.
2. Vogel A, Meyer T, Sapisochin G, Salem R, Saborowski A. Hepatocellular carcinoma. *Lancet* 2022;400:1345-1362.
3. Berzigotti A, Reig M, Abraldes JG, Bosch J, Bruix J. Portal hypertension and the outcome of surgery for hepatocellular carcinoma in compensated cirrhosis: a systematic review and meta-analysis. *Hepatology* 2015;61:526-536.
4. Johnson PJ, Berhane S, Kagebayashi C, Satomura S, Teng M, Reeves HL, et al. Assessment of liver function in patients with hepatocellular carcinoma: a new evidence-based approach-the ALBI grade. *J Clin Oncol* 2015;33:550-558.
5. Demirtas CO, D'Alessio A, Rimassa L, Sharma R, Pinato DJ. ALBI grade: Evidence for an improved model for liver functional estimation in patients with hepatocellular carcinoma. *JHEP Rep* 2021;3:100347.
6. Wong GL, Hui VW, Tan Q, Xu J, Lee HW, Yip TC, et al. Novel machine learning models outperform risk scores in predicting hepatocellular carcinoma in patients with chronic viral hepatitis. *JHEP Rep* 2022;4:100441.
7. Ho CT, Tan EC, Lee PC, Chu CJ, Huang YH, Huo TI, et al. Conventional and machine-learning based risk score for patients with early-stage hepatocellular carcinoma. *Clin Mol Hepatol* 2024.
8. Lin H, Li G, Delamarre A, Ahn SH, Zhang X, Kim BK, et al. A Liver Stiffness-Based Etiology-Independent Machine Learning Algorithm to Predict Hepatocellular Carcinoma. *Clin Gastroenterol Hepatol* 2024;22:602-610 e607.
9. Singal AG, Llovet JM, Yarrow M, Mehta N, Heimbach JK, Dawson LA, et al. AASLD Practice Guidance on prevention, diagnosis, and treatment of hepatocellular carcinoma. *Hepatology* 2023;78:1922-1965.
10. Li X, Huang X, Lei L, Tong S. Impact of sarcopenia and sarcopenic obesity on survival in patients with primary liver cancer: a systematic review and meta-analysis. *Front Nutr* 2023;10:1233973.
11. Chang K-V, Chen J-D, Wu W-T, Huang K-C, Hsu C-T, Han D-S. Association between Loss of Skeletal Muscle Mass and Mortality and Tumor Recurrence in Hepatocellular Carcinoma: A Systematic Review and Meta-Analysis. *Liver Cancer* 2018;7.
12. Chan SL, Wong LL, Chan KA, Chow C, Tong JH, Yip TC, et al. Development of a Novel Inflammation-Based Index for Hepatocellular Carcinoma. *Liver Cancer* 2020;9:167-181.
13. Yip TC, Hui VW, Tse YK, Wong GL. Statistical strategies for HCC risk prediction models in patients with chronic hepatitis B. *Hepatoma Research* 2021;7:7.
14. Zuo R, Li G, Choi B, S Bhowmick S, Mah DN-y, Wong GLH. SVP-T: A Shape-Level Variable-Position Transformer for Multivariate Time Series Classification. *Proceedings of the AAAI Conference on Artificial Intelligence* 2023;37:11497-11505.
15. Tan Q, Ye M, Ma AJ, Yang B, Yip TCF, Wong GLH, et al. Explainable Uncertainty-Aware Convolutional Recurrent Neural Network for Irregular Medical Time Series. *IEEE Transactions on Neural Networks and Learning Systems* 2021;32:4665-4679.
16. He Y, Hu B, Zhu C, Xu W, Ge Y, Hao X, et al. A Novel Multimodal Radiomics Model for Predicting Prognosis of Resected Hepatocellular Carcinoma. *Front Oncol* 2022;12:745258.
17. Yip TC, Lee HW, Chan WK, Wong GL, Wong VW. Asian perspective on NAFLD-associated HCC. *J Hepatol* 2022;76:726-734.

Figure 1. Prediction of prognosis of patients with early hepatocellular carcinoma using machine learning models.



Accepted article