

LETTER TO THE EDITOR

Development and validation of an artificial intelligence model for predicting post-transplant hepatocellular cancer recurrence

Dear Editor,

In recent years, criteria based on the combination of morphology and biology have been proposed for improving the selection of hepatocellular cancer (HCC) patients waiting for liver transplantation (LT) [1, 2]. Since all the proposed models showed suboptimal results in predicting the risk of post-LT recurrence, a prediction model constructed using artificial intelligence (AI) could be an attractive way to surpass this limit [3, 4]. Therefore, the Time_Radiological-response_Alpha-fetoproteIN_Artificial-Intelligence (TRAIN-AI) model was developed, combining morphology and biology tumor variables.

A Training Set ($n = 2,936$) derived from an International Cohort was adopted to create the model. A Validation Set ($n = 734$) derived from the same International Cohort and an external Test Set ($n = 356$) were identified for internal and external validation of TRAIN-AI, respectively (Supplementary Figure S1). Training and Validation Sets presented similar characteristics (Supplementary Table S1). Conversely, relevant differences were observed when the Test Set was compared with the Validation Set; therefore, external validation of the model was performed in a very different population (i.e., Test Set) from the one from which the TRAIN-AI was derived and internally tested (i.e., Training and Validation Sets) (Supplementary Table S2).

1 | TRAIN-AI MODEL VARIABLES

Eight variables were significantly associated with the risk of recurrence and used for constructing the TRAIN-AI

model: target lesion diameter, nodules number, alpha-fetoprotein, waiting time length, radiological response, model for end-stage liver disease (MELD), living donor liver transplantation, and center volume (Supplementary Table S3). The statistical approaches used for constructing the model are reported in the [Supplementary Material](#).

The average impact of each factor on the model output magnitude was explored, with the nodules number and the alpha-fetoprotein (AFP) identified as the most relevant variables (Supplementary Figure S2).

2 | INTERNAL VALIDATION (VALIDATION SET)

Table 1 summarizes the accuracy of the TRAIN-AI model when compared to several currently adopted criteria for predicting post-LT HCC recurrence [1, 5-7].

The internal validation was performed using the Validation Set data. Time-dependent concordance by Antolini et al. [8] showed that the TRAIN-AI model had the best accuracy (concordance = 0.77; 95% confidence interval [CI] = 0.72-0.82). The TRAIN-AI model consistently outperformed the other criteria (AFP-French model concordance = 0.68; Metroticket 2.0 = 0.68; Milan Criteria [MC] = 0.63) (Table 1).

To clarify the magnitude of prediction improvement obtained using the TRAIN-AI score, the Brier score and the Brier skill score were calculated. The TRAIN-AI reported the best value (Brier score = 0.10) among the different criteria. Comparing the TRAIN-AI with each other score, an improvement of the prediction was observed in all the cases: the best progress was reported by comparing the TRAIN-AI score with MC (Brier Skill Score + 14.26%) (Table 1).

TRAIN-AI also had the best Harrell c-statistics for the 5-year recurrence risk (concordance = 0.77, 95% CI = 0.71-0.82), being markedly superior to the other criteria

Abbreviations: AFP, alpha-fetoprotein; AI, artificial intelligence; CI, confidence interval; HCC, hepatocellular cancer; LT, liver transplantation; MC, Milan Criteria; MELD, model for end-stage liver disease; TRAIN-AI, Time_Radiological-response_Alpha-fetoproteIN_Artificial-Intelligence.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](#) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *Cancer Communications* published by John Wiley & Sons Australia, Ltd. on behalf of Sun Yat-sen University Cancer Center.

TABLE 1 Accuracy of TRAIN-AI compared to currently widely used criteria for post-LT HCC recurrence: Validation Set (internal validation) and Test Set (external validation).

Criteria	Time-dependent concordance (95% CI)*	Brier score	Brier skill score (%)**	Harrell c-statistics (5-year recurrence) (95% CI)	P
Validation Set (internal validation)					
TRAIN-AI	0.77 (0.72-0.82)	0.10	Ref.	0.77 (0.71-0.82)	Ref.
AFP-French model	0.68 (0.64-0.73)	0.15	5.09	0.67 (0.60-0.74)	< 0.001
Metroticket 2.0 score	0.68 (0.63-0.73)	0.18	8.19	0.68 (0.61-0.74)	< 0.001
MC	0.63 (0.58-0.68)	0.23	14.26	0.64 (0.58-0.71)	< 0.001
San Francisco criteria	0.61 (0.57-0.66)	0.20	10.74	0.62 (0.55-0.69)	< 0.001
Up-to-Seven criteria	0.61 (0.56-0.66)	0.19	9.74	0.62 (0.55-0.69)	< 0.001
Asan criteria	0.59 (0.54-0.63)	0.16	6.65	0.59 (0.52-0.65)	< 0.001
Kyoto criteria	0.57 (0.53-0.61)	0.15	5.37	0.57 (0.50-0.64)	< 0.001
HALT-HCC score	0.53 (0.51-0.55)	0.12	2.84	0.53 (0.46-0.59)	< 0.001
Test Set (external validation)					
TRAIN-AI	0.77 (0.70-0.84)	0.10	Ref.	0.78 (0.71-0.85)	Ref.
Metroticket 2.0 score	0.69 (0.63-0.76)	0.18	7.34	0.69 (0.59-0.78)	0.020
AFP-French model	0.67 (0.61-0.75)	0.17	6.23	0.66 (0.56-0.75)	0.006
MC	0.66 (0.58-0.73)	0.24	13.94	0.67 (0.58-0.76)	0.007
Kyoto criteria	0.65 (0.58-0.72)	0.18	7.34	0.64 (0.55-0.73)	0.002
Asan criteria	0.65 (0.57-0.72)	0.18	7.61	0.64 (0.55-0.73)	0.002
San Francisco criteria	0.64 (0.58-0.72)	0.19	8.43	0.65 (0.56-0.74)	0.004
Up-to-Seven criteria	0.62 (0.55-0.70)	0.19	8.98	0.63 (0.54-0.72)	0.001
HALT-HCC score	0.52 (0.50-0.55)	0.15	4.58	0.52 (0.43-0.61)	< 0.001

Abbreviations: CI, confidence intervals; TRAIN, Time Radiological response Alpha-fetoprotein; AI, artificial intelligence; AFP, alpha-fetoprotein; MC, Milan Criteria; HALT-HCC, Hazard Associated with Liver Transplantation for Hepatocellular Carcinoma.

Note: The criteria composed of continuous values were not dichotomized.

*All the reported time-dependent concordance values and 95% CI are means calculated after a 1,000-fold bootstrap method. The concordance was estimated using the time-dependent concordance analysis by Antolini et al. [8].

**The reported values of the Brier skill scores correspond to the percentage of prediction improvement of the TRAIN-AI when compared with other criteria.

(AFP-French model = 0.67, $P < 0.001$; Metroticket 2.0 = 0.68, $P < 0.001$; MC = 0.64, $P < 0.001$) (Table 1). Sub-analyses confirmed the prognostic ability of the TRAIN-AI also in the setting of hepatitis C or Hepatitis B viruses - positivity, LT performed in Asia or Europe, or exceeding the MC status (Supplementary Table S4).

3 | EXTERNAL VALIDATION (TEST SET)

Also, in the Test Set data, the TRAIN-AI model had the best concordance (concordance = 0.77; 95% CI = 0.70-0.84). The TRAIN-AI model consistently outperformed the other criteria (Metroticket 2.0 = 0.69; AFP-French model = 0.67; MC = 0.66) (Table 1).

The TRAIN-AI Brier score showed the best value (Brier score = 0.10) among the different criteria. Comparing the TRAIN-AI with each other score, an improvement of the

prediction was observed in all the cases: the best progress was reported with the MC (Brier Skill Score + 13.94%) (Table 1).

The TRAIN-AI c-statistics for the risk of 5-year recurrence was the best observed (concordance = 0.78, 95% CI = 0.71-0.85), being markedly superior to the other criteria (Metroticket 2.0 = 0.69, $P = 0.020$; MC = 0.67, $P = 0.007$; AFP-French model = 0.66, $P = 0.006$) (Table 1).

4 | CALIBRATION OF THE MODEL IN INDIVIDUAL PATIENTS

A model user-friendly web calculator was constructed (<https://train-ai.cloud>) and made available for calculating the expected recurrence after LT in individual patients.

After the stratification of the explored populations in three 5-year recurrence risk classes (low: $\leq 15\%$;

intermediate:16%-30%; high: > 30%), the expected vs. observed recurrence rates were compared in the Validation and Test Sets (Supplementary Figure S3).

Starting from the assumption that the Hosmer-Lemeshow test indicates a poor calibration if $P < 0.050$, the test showed a good calibration in the Validation Set ($P = 0.540$) and in the Test Set ($P = 0.380$) (Supplementary Figure S3).

5 | IMPLICATIONS OF USING THE MODEL

This is the largest prediction model published in this field based on deep learning algorithms. The performances of TRAIN-AI outperformed several currently used HCC selection criteria both in the internal and external validation. A user-friendly web calculator was also created to calculate each patient's recurrence risk.

The proposed model is based only on well-recognized variables readily available worldwide, consenting to reach high standardization rates, completeness, and granularity.

Another relevant aspect of this AI model is that it can continuously evolve with further data accumulation. The web calculator allows TRAIN-AI to improve its prognostic performance through continuous data training enlargement. To consent to this improvement, two collaborative international consortia routinely updating their data (i.e., the EurHeCaLT and the East-West LT Study Groups) have been involved in this project.

Recently, two studies focused on post-LT HCC recurrence based on AI models [3, 4]. The main disadvantage of these studies was the limited number of patients available for model development and training. Deep learning models typically require thousands of data. This shortcoming is not present in our study, in which 2,936 patients were used for constructing the Training Set.

Another relevant problem was the prediction “overfitting” phenomenon, which may generate overly optimistic results [9]. This problem is relevant when training and validation sets derive from the same population. To solve this limit, we externally tested the model using a geographically different population. Training and Validation Sets were composed of Euro-Asiatic patients with short waiting times, one-third of living donation cases, and three-quarter of cases with neo-adjuvant therapies. Conversely, the Test Set was based on North-American patients with long waiting times, fewer cases of living donation, and almost all the cases treated with neo-adjuvant therapies. Despite these differences, the concordance of the TRAIN-AI was always very good (0.77 in both Validation and Test Sets) (Table 1),

with a percentage of prediction improvement markedly encompassing all the other criteria.

6 | LIMITS OF THE STUDY

This presented study has some limits. First, it is impossible to understand the outcome operations resulting from deep learning. Secondly, the study is retrospective. Thirdly, some variables were not used for the TRAIN-AI construction, like des-gamma carboxy-prothrombin, inflammatory markers, radiologically detectable macrovascular invasion, and radiomics [10].

7 | CONCLUSION

The TRAIN-AI model showed higher accuracy than other frequently used scores for the risk of post-LT HCC recurrence. A user-friendly web calculator has been developed to improve the model's availability. A tailored and justified transplantability cutoff can be proposed stratifying the patients in recurrence risk classes. A further prediction implementation of the AI model can be obtained by increasing the number of patients for training.

#Collaborators of the EurHeCaLT and West-East LT Collaborative Effort Study Groups

Austria: Andre Viveiros (University of Innsbruck, Innsbruck); *Belgium:* Samuele Iesari (Université Catholique de Louvain, Brussels), Olga Ciccarelli (UCL, Brussels); *Croatia:* Branislav Kocman (University of Zagreb, Zagreb); *Germany:* Jens Mittler (University of Mainz, Mainz); *Hong Kong:* Tiffany Wong (University of Hong Kong, Hong Kong); *India:* Arvinder Singh Soin (Medanta-The Medicity, Gurgaon); *Italy:* Federico Mocchegiani (Polytechnic University of Marche, Ancona), Matteo Cescon (University of Bologna, Bologna), Alessandro Vitale (University of Padua, Padua), Gianluca Mennini (Sapienza University, Rome), Tommaso Maria Manzia (PTV University, Rome), Alfonso W. Avolio (Catholic University, Rome), Gabriele Spoletini (Catholic University, Rome), Marco Colasanti (San Camillo Hospital, Rome); *Japan:* Tomoharu Yoshizumi (Kyushu University, Fukuoka), Toshimi Kaido, Etsuro Hatano (Graduate School of Medicine, Kyoto); *Taiwan:* Chih Che Lin (Kaohsiung, Taiwan); *United Kingdom:* Margarita Papatheodoridi (Royal Free Hospital, London), Simona Onali (Royal Free Hospital, London); *United*

States of America: Karim Halazun (Columbia University, New York).

DECLARATIONS

AUTHOR CONTRIBUTIONS

Quirino Lai and Carmine De Stefano contributed to the conception and design of the study; Quirino Lai, Prashant Bhangui, Toru Ikegami, Benedikt Schaefer, Maria Hoppe-Lotichius, Anna Mrzljak, Takashi Ito, Marco Vivarelli, Giuseppe Tisone, Salvatore Agnes, Giuseppe Maria Ettore, Massimo Rossi, Emmanuel Tsochatzis, Chung Mau Lo, Chao-Long Chen, Umberto Cillo, Matteo Ravaioli, and Jan Paul Lerut contributed to acquisition of data; Quirino Lai and Carmine De Stefano analyzed and interpreted the data; Quirino Lai, Carmine De Stefano and Jan Paul Lerut drafted the article; Jean Emond, Toru Ikegami, Benedikt Schaefer, Maria Hoppe-Lotichius, Marco Vivarelli, Emmanuel Tsochatzis, and Matteo Ravaioli critically revised the manuscript; and all authors approved the final version.

ACKNOWLEDGMENT

None.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to declare about the present study.

FUNDING

The authors have not received any support for the present study, and no specific funding was used for this study.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

The study was performed according to the Declaration of Helsinki. The study was approved by the Umberto I Policlinico of Rome Institutional Review Board (Approval number: 1000/2018).

CONSENT FOR PUBLICATION

Not applicable

DATA AVAILABILITY STATEMENT

Individual, de-identified patient data and data dictionary can be made available at the request of investigators who propose to use the data in a way that has been approved by all the members of the Study Group following a review of a methodologically sound research proposal. Data will be made available 6 months after article publication, with no end date. Requests for de-identified data should be made to the study Chief Investigator (Quirino Lai).

Quirino Lai¹ 
 Carmine De Stefano²
 Jean Emond³
 Prashant Bhangui⁴
 Toru Ikegami⁵
 Benedikt Schaefer⁶
 Maria Hoppe-Lotichius⁷
 Anna Mrzljak⁸
 Takashi Ito⁹
 Marco Vivarelli¹⁰
 Giuseppe Tisone¹¹
 Salvatore Agnes¹²
 Giuseppe Maria Ettore¹³
 Massimo Rossi¹
 Emmanuel Tsochatzis¹⁴
 Chung Mau Lo¹⁵
 Chao-Long Chen¹⁶
 Umberto Cillo¹⁷
 Matteo Ravaioli¹⁸
 Jan Paul Lerut¹⁹

the EurHeCaLT and the West-East LT Study Group^{1, #}

¹*General Surgery and Organ Transplantation Unit, AOU Policlinico Umberto I, Sapienza University of Rome, Rome, Italy*

²*Data Science and Engineering, Polytechnic of Turin, Turin, Italy*

³*Division of Liver Transplantation and Hepatobiliary Surgery, Department of Surgery, Weill Cornell Medicine-Columbia University, New York, US*

⁴*Medanta Institute of Liver Transplantation and Regenerative Medicine, Medanta-The Medicity, Gurgaon, India*

⁵*Department of Surgery and Science, Kyushu University, Fukuoka, Japan*

⁶*Department of Medicine I, Gastroenterology, Hepatology and Endocrinology, Medical University of Innsbruck, Innsbruck, Austria*

⁷*Klinik für Allgemein-, Viszeral- und Transplantationschirurgie, Universitätsmedizin Mainz, Mainz, Germany*

⁸*Liver Transplant Centre, Merkur University of Zagreb, Zagreb, Croatia*

⁹*Division of Hepato-Biliary-Pancreatic and Transplant Surgery, Department of Surgery, Graduate School of Medicine, Kyoto, Japan*

¹⁰*Unit of Hepatobiliary Surgery and Transplantation, AOU Ospedali Riuniti, Polytechnic University of Marche, Ancona, Italy*

¹¹*Department of Surgical Sciences and Medical Sciences University of Rome-Tor Vergata, Rome, Italy*

¹²Liver Unit, Department of Surgery, Catholic University-Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Rome, Italy

¹³Department of Transplantation and General Surgery, San Camillo Hospital, Rome, Italy

¹⁴UCL Institute for Liver and Digestive Health and Royal Free Sheila Sherlock Liver Centre, Royal Free Hospital, London, UK

¹⁵Hong Kong University-Department of Surgery, Queen Mary Hospital, University of Hong Kong, Hong Kong, P. R. China

¹⁶Department of Surgery, Kaohsiung Chang Gung Memorial Hospital, Chang Gung University College of Medicine, Kaohsiung, Taiwan, P. R. China

¹⁷Department of Surgery, Oncology and Gastroenterology, University of Padua, Padua, Italy

¹⁸Department of General Surgery and Transplantation, IRCCS, Azienda Ospedaliero-Universitaria di Bologna, Bologna University, Bologna, Italy

¹⁹Institut de Recherche Clinique, Université catholique de Louvain, Brussels, Belgium

Correspondence

Quirino Lai, MD Ph.D., General Surgery and Organ Transplantation Unit, Department of General and Specialistic Surgery, Sapienza University of Rome, AOU Policlinico Umberto I of Rome, Viale del Policlinico 155, 00161 Rome, Italy.

Email: quirino.lai@uniroma1.it

#Collaborators are listed at the end of the article

ORCID

Quirino Lai  <https://orcid.org/0000-0003-1487-3235>

REFERENCES

1. Firl DJ, Sasaki K, Agopian VG, Gorgen A, Kimura S, Dumronggittigule W, et al. European Hepatocellular Cancer Liver Transplant Study Group. Aucejo FN. Charting the

path forward for risk prediction in liver transplant for hepatocellular carcinoma: International validation of HALTHCC among 4,089 patients. *Hepatology*. 2020;71(2):569–82.

2. Lai Q, Nicolini D, Inostroza Nunez M, Iesari S, Goffette P, Agostini A, et al. A Novel Prognostic Index in Patients With Hepatocellular Cancer Waiting for Liver Transplantation: Time-Radiological-response-Alpha-fetoprotein-INflammation (TRAIN) Score. *Ann Surg*. 2016;264(5):787–96.
3. Ivanics T, Nelson W, Patel MS, Claasen MPAW, Lau L, Gorgen A, et al. The Toronto Postliver Transplantation Hepatocellular Carcinoma Recurrence Calculator: A Machine Learning Approach. *Liver Transpl*. 2022;28(4):593–602.
4. Nam JY, Lee JH, Bae J, Chang Y, Cho Y, Sinn DH, et al. Novel Model to Predict HCC Recurrence after Liver Transplantation Obtained Using Deep Learning: A Multicenter Study. *Cancers (Basel)*. 2020;12(10):2791.
5. Mazzaferro V, Regalia E, Doci R, Andreola S, Pulvirenti A, Bozzetti F, et al. Liver transplantation for the treatment of small hepatocellular carcinomas in patients with cirrhosis. *N Engl J Med*. 1996;334(11):693–99.
6. Duvoux C, Roudot-Thoraval F, Decaens T, Pessione F, Badran H, Piardi T, et al. Liver Transplantation French Study Group. Liver transplantation for hepatocellular carcinoma: a model including α -fetoprotein improves the performance of Milan criteria. *Gastroenterology*. 2012;143(4):986–94.e3.
7. Mazzaferro V, Sposito C, Zhou J, Pinna AD, De Carlis L, Fan J, et al. Metroticket 2.0 Model for Analysis of Competing Risks of Death After Liver Transplantation for Hepatocellular Carcinoma. *Gastroenterology*. 2018;154(1):128–39.
8. Antolini L, Boracchi P, Biganzoli E. A time-dependent discrimination index for survival data. *Stat Med*. 2005;24(24):3927–44
9. Obermeyer Z, Emanuel EJ. Predicting the future - Big data, machine learning, and clinical medicine. *N Engl J Med*. 2016;375(13):1216–9.
10. Assalino M, Terraz S, Grat M, Lai Q, Vachharajani N, Gringeri E, et al. Liver transplantation for hepatocellular carcinoma after successful treatment of macrovascular invasion - a multi-center retrospective cohort study. *Transpl Int*. 2020;33(5):567–75.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.