MoRAL-AI: AI-based Liver Transplantation Model

Subjects: Gastroenterology & Hepatology | Computer Science, Artificial Intelligence

Contributor: Joon Yeul Nam , Jeong-Hoon Lee , Junho Bae , Young Chang , Yuri Cho , Dong Hyun Sinn , Bo Hyun Kim , Seoung Hoon Kim , Nam-Joon Yi , Kwang-Woong Lee , Jong Man Kim , Joong-Won Park , Yoon Jun Kim , Jung-Hwan Yoon , Jae-Won Joh , Kyung-Suk Suh

Novel model to predict HCC recurrence after liver transplantation based on deep learning

hepatocellular carcinoma   liver transplantation   deep learning   prediction model

1. Introduction

Liver transplantation (LT) can be the most effective treatment among all treatment options for hepatocellular carcinoma (HCC) in carefully selected patients who meet certain criteria. Although the Milan criteria (MC) were introduced in 1996, they are still the most widely used system[1]. While the MC have been extensively used, recent data suggest that the MC may be too conservative when selecting an LT candidate[2]. Therefore, several other models have been developed, such as the University of California San Francisco (UCSF), up-to-seven, and the Kyoto criteria and model, to predict tumor recurrence after living donor LT (MoRAL) scores[3][4][5][6]. However, there are still no standard criteria. Moreover, due to continuous advances in diagnostic and therapeutic techniques for HCC, the establishment of standard criteria for LT becomes increasingly difficult.

Several factors related to HCC recurrence were identified in previous models, including maximum tumor diameter, tumor number, portal vein invasion, and serum tumor markers such as alpha-fetoprotein (AFP) and protein induced by vitamin K absence-II (PIVKA-II). When these factors were applied to previous models, a cutoff value of each factor was determined by conventional statistical methods. However, each factor was applied to the models as binary data, despite comprising continuous values. Models relying on binary data are simple and intuitive but may have inferior accuracy to ones built with continuous data. Moreover, the results of each previous model were also dichotomous: HCC recurrence or no recurrence after LT. It would be more helpful for precise medical decisions if the risk of HCC recurrence was presented as a continuous recurrence probability according to follow-up duration rather than as a simple dichotomous conclusion.

2. MoRAL-AI for tumor recurrence after LT

Authors developed and validated a novel prediction model, called MoRAL-AI, for tumor recurrence after LT in patients with HCC. To the best of our knowledge, this is the first prediction model based on deep learning algorithms. The performance of MoRAL-AI was confirmed by using an independent validation cohort and was

better than that of the MC, currently the most widely used criteria, as well as other prediction models. The MoRAL-AI is served through the website, and it can evolve with further data accumulation from various cohort groups. With this evolution, More evolved criteria of LT could be established.

For LT as a treatment option for HCC, the underlying hepatic function of the recipient is not an important factor. While the recipient usually has liver cirrhosis before LT, the recipient liver is completely replaced by the donor liver and the severity of the pre-LT cirrhosis might thus not affect the post-LT tumor recurrence[7]. Therefore, predictive factors that are related to the post-LT HCC recurrence might include only the tumor burden and the biological aggressiveness of tumor cells before LT. Imaging studies can provide tumor-related information (e.g., tumor number, maximum tumor diameter, and vascular tumor invasion). On the other hand, serum levels of tumor markers (AFP and PIVKA-II) can reflect both tumor burden and biological aggressiveness because AFP and PIVKA-II had a significantly positive correlation with histological aggressive findings (microvascular invasion, perineural invasion, and serosal invasion) as well as tumor burden. Among previous models, the MC, UCSF, and up-to-seven models comprise only the factors related to imaging-based tumor burden, whereas the MoRAL score consists of only serum tumor markers[1][3][5][6]. The Hangzhou criteria and Kyoto criteria consider both imaging-based tumor burden and tumor markers[4][8]. The MoRAL-AI was developed based on both imaging-based tumor burden and biochemical tumor markers to maximize its performance.

The general disadvantage of the deep learning method is that it typically requires a large amount of data; it has previously been applied to the analysis of medical images such as plain X-ray, CT, or histology images, whose data can be abundantly obtained since a number of images are being taken during daily clinical practice[9][10]. However, when a specific disease is being analyzed, the size of the available data set is generally limited. For example, in Korea, only 1400 cases of LT were performed in 2015[11], whereas approximately 265,000,000 plain X-rays were performed[12]. Moreover, it is generally complicated to collect demographic information and clinical results via a unified form due to the different data forms used by hospitals. Thus, data from fewer than 1000 cases have been analyzed with conventional statistical methods that can evaluate a relatively small dataset. However, our current model was derived from a relatively small derivation cohort containing 349 patients and showed a better predictive power than previous models. This result might suggest that DNN models can be developed even with relatively few data points if the potential prognostic factors for the prediction model have previously been well identified. Indeed, the predictive factors used in this study, such as tumor diameter[1][3][4][5][13], tumor number [1][3][4][5], AFP[6][13], PIVKA-II[4][6], and portal vein invasion[1][3][4][5][13][14], were well-identified parameters in previous studies.

The DNN model has several strengths compared with conventional statistical modeling. First, by applying a deep learning method to the prediction model, it is possible to derive more accurate continuous probability results. In the conventional statistical method, a scoring system is established to divide risk groups and continuous variables are stratified according to arbitrary cutoffs. However, DNN models can use continuous data rather than converting them into categorical or binary variables. Thus, they can provide more accurate and individualized results, which means that they can calculate the individual tumor recurrence probability at any time point after an LT according to baseline clinical information. Second, because DNN models were originally designed to be calculated by computer, there is no need to focus on ease of use. In contrast, previous conventional models had to be intuitive and easy to

use and simple models comprising fewer factors were thus preferred. However, because DNN models involve an automatic calculation on a web application, there is no need to limit the number of factors or to consider the complexity of the formulae. Third, DNN models can evolve continuously with further data accumulation. Previous models, such as the MC system which was developed in 1996, have not been changed, despite the accumulation of new data. However, DNN models can continuously improve their performances through additional data training.

This study has several limitations. First, it is impossible to understand the outcome operations resulting from deep learning. This is a general shortcoming of deep learning methods. Second, because this model was developed from Asian patients who underwent living donor LTs and whose underlying liver diseases is predominantly chronic hepatitis B, further validation in Western countries and deceased donor LT cohorts is warranted. Our model can provide additional options to select for deceased or living donor LT in our web application with further validation. Third, PIVKA-II is generally measured in Asian countries before LT, but is less commonly measured in Western countries. With further data training with other data sets, it may be possible to develop another model with high performance power without a certain factor like PIVKA-II.

References

1. Mazzaferro, V.; Regalia, E.; Doci, R.; Andreola, S.; Pulvirenti, A.; Bozzetti, F.; Montalto, F.; Ammatuna, M.; Morabito, A.; Gennari, L. Liver transplantation for the treatment of small hepatocellular carcinomas in patients with cirrhosis. N. Engl. J. Med. 1996, 334, 693–700.

2. Herrero, J.I.; Sangro, B.; Pardo, F.; Quiroga, J.; Inarrairaegui, M.; Rotellar, F.; Montiel, C.; Alegre, F.; Prieto, J. Liver transplantation in patients with hepatocellular carcinoma across Milan criteria. Liver Transplant. 2008, 14, 272–278.

3. Yao, F.Y.; Ferrell, L.; Bass, N.M.; Watson, J.J.; Bacchetti, P.; Venook, A.; Ascher, N.L.; Roberts, J.P. Liver transplantation for hepatocellular carcinoma: Expansion of the tumor size limits does not adversely impact survival. Hepatology 2001, 33, 1394–1403.

4. Ito, T.; Takada, Y.; Ueda, M.; Haga, H.; Maetani, Y.; Oike, F.; Ogawa, K.; Sakamoto, S.; Ogura, Y.; Egawa, H. Expansion of selection criteria for patients with hepatocellular carcinoma in living donor liver transplantation. Liver Transplant. 2007, 13, 1637–1644.

5. Mazzaferro, V.; Llovet, J.M.; Miceli, R.; Bhoori, S.; Schiavo, M.; Mariani, L.; Camerini, T.; Roayaie, S.; Schwartz, M.E.; Grazi, G.L. Predicting survival after liver transplantation in patients with hepatocellular carcinoma beyond the Milan criteria: A retrospective, exploratory analysis. Lancet Oncol. 2009, 10, 35–43.

6. Lee, J.-H.; Cho, Y.; Kim, H.Y.; Cho, E.J.; Lee, D.H.; Yu, S.J.; Lee, J.W.; Yi, N.-J.; Lee, K.-W.; Kim, S.H. Serum tumor markers provide refined prognostication in selecting liver transplantation

candidate for hepatocellular carcinoma patients beyond the Milan criteria. Ann. Surg. 2016, 263, 842–850.

7. Schwartz, M. Liver transplantation for hepatocellular carcinoma. Gastroenterology 2004, 127, S268–S276.

8. Xu, X.; Lu, D.; Ling, Q.; Wei, X.; Wu, J.; Zhou, L.; Yan, S.; Wu, L.; Geng, L.; Ke, Q. Liver transplantation for hepatocellular carcinoma beyond the Milan criteria. Gut 2016, 65, 1035–1041.

9. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; Van Der Laak, J.A.; Van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. Med. Image Anal. 2017, 42, 60–88.

10. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. Nature 2015, 521, 436.

11. Korean Network for Organ Sharing. Available online: https://www.konos.go.kr/konosis/common/bizlogic.jsp or https://www.konos.go.kr; (accessed on 8 April 2020).

12. Ministry of the Interior and Safety. Available online: http://www.prism.go.kr/homepage/entire/retrieveEntireDetail.do?research_id=1351000-201700153 or http://www.prism.go.kr; (accessed on 8 April 2020).

13. Lee, K.W.; Suh, S.W.; Choi, Y.; Jeong, J.; Yi, N.J.; Kim, H.; Yoon, K.C.; Hong, S.K.; Kim, H.S.; Lee, K.B.J.L.T. Macrovascular invasion is not an absolute contraindication for living donor liver transplantation. Liver Transpl. 2017, 23, 19–27.

14. Schmidhuber, J. Deep learning in neural networks: An overview. Neural Netw. 2015, 61, 85–117.