90057_Auto_Edited.docx

Name of Journal: *World Journal of Clinical Cases*
Manuscript NO: 90057
Manuscript Type: LETTER TO THE EDITOR

Machine learning in liver surgery: benefits and pitfalls

Machine Learning in liver surgery

Rafael Calleja, Manuel Durán, María Dolores Ayllón, Ruben Ciria, Javier Briceño

Abstract

The application of machine learning algorithms in various fields of hepatology is an issue of interest. However, we must to be cautious with the results. In this letter, based on a published machine learning prediction model for acute kidney injury after liver surgery, we discuss some limitations of machine learning models and how they may be addressed in the future. Although the future faces significant challenges, it also holds a great potential.

Key Words: Machine learning; Liver surgery; Artificial intelligence; Random Forest; Prediction model

Calleja R, Durán M, Ayllón MD, Ciria R, Briceño J. Machine learning in liver surgery: benefits and pitfalls. *World J Clin Cases* 2024; In press

Core Tip: Artificial intelligence is trending topic in healthcare research. Machine learning classifiers have been explored in the field of liver surgery and liver transplantation. However, despite of promising results, a real applicability is limited by several factors.

TO THE EDITOR

We read with interest a retrospective study by Dong JF *et al*.[1], who developed a machine learning (ML) prediction model for acute kidney injury (AKI) following liver resection. We congratulate the authors for their work and contribution to this field. Liver resection (LR) is the treatment of a wide spectrum of liver lesions. However, after LR, the reported incidence of AKI post-LR ranges from 10 to 15%[2], significantly impacting patient morbidity and mortality. Hence, identifying factors that may lead to the development of this condition is relevant. In this regard, Dong JF *et al*. [1] have explored the potential contribution of ML classifiers to this issue.

The authors analyzed a retrospective c0h0rt 0f 2450 patients, training and validating four ML classifiers (Logistic regression, Random Forest, Support vector machine, Extreme gradient boosting and Decision Tree). The training methodology (10-fold cross-validation) and validation (a hold-out technique with 30% of patterns) were deemed adequate. Finally, Random-Forest exhibited the highest performance (AUC= 0.92) among the classifiers. While the achieved results were satisfactory, certain considerations must be addressed.

Firstly, the rate of missing values should be reported. A significant proportion of missing values can impact model training, subsequently affecting model performance and generalizability. Hence, RF classifiers are the best algorithm to deal with a significant rate of missing values[3]. Conversely, if this rate is low, artificial neural networks (ANNs) could offer promising results in the dataset. Secondly, several factors reported in the literature are associated with AKI after LR such as major hepatectomy, operation duration, hepatojejunostomy, increased MELD or blood-transfusion[2, 4-8]. Among these factors, only surgery time has been found in baseline characteristics. To consider including these variables may increase model robustness. Finally, to perform an external validation is a challenge. Differences between the training and external validation cohorts may affect model accuracy. In this regard a prospective validation could be an alternative.

Some of the latest research in ML applications ranges from protein structure prediction or Covid-19 diagnosis from X-ray images, to optimizing donor-recipient matching to reduce waitlist mortality or improve post-transplant Outcomes[9-11]. Our experience in the field of ML in liver surgery stems from liver transplantation. Our efforts have primarily focused on improving donor-recipient matching. Based on graft survival as end-point, we developed an ANN-model that achieved a performance, in terms of area under the curve (AUC), of around 0.82[12]. This methodology was validated in an external cohort, improving AUC by up to 15%[13]. This ANN was integrated into a rule system along the

MELD score to prioritize graft allocation. Although this methodology was explored in the UNOS database, limited results due to a significant proportion of missing values were found[14]. Dong JF *et al*.[1] found that their model performance was better than the current scores for AKI prediction. Similarly, we reported how different ML models outperformed traditional scores such as MELD, SOFT, DRI and BAR (**Figure 1**). In medicine there are variables that do not necessarily have to assume a linear relationship. Hence, ML models are superior to statistical methods (such as linear regression), from which most of these scores are derived. [15] However, these findings may be attributed to model overtraining, so a validation is required.

One of the biggest lessons learned from the use of these models is their high dependency on the dataset on which they are trained. This issue affects their real applicability. Retrospective data, external validation, the “black box issue” in ANN or data protection policies are other significant contributing factors. To overcome these barriers, better data handling policies are needed. The applicability relies on the confidence of clinicians in the use of these models. In this regard, if external validations are not possible (we advocate region-specific rather than universal models), prospective validations should be considered. Moreover, databases must be regularly updated to reinforce the learning of these models. Clinical scenarios are dynamics, and models must learn and change with them.

Interest in artificial intelligence and ML has increased in recent years. They are able to handle large amounts of data in a fast way, yielding accurate results. However, we must be aware of the limitations of these models and address them to achieve a real integration.

90057_Auto_Edited.docx

ORIGINALITY REPORT

0%
SIMILARITY INDEX

PRIMARY SOURCES

EXCLUDE QUOTES	ON	EXCLUDE SOURCES	< 12 WORDS
EXCLUDE BIBLIOGRAPHY	ON	EXCLUDE MATCHES	< 12 WORDS