Editorial

Outcome prediction after resection of colorectal cancer liver metastases: out with the old, in with the new?

Iakovos Amygdalos[1]^, Daniel Truhn[2], Florian W. R. Vondran[1]

[1]Department of General, Visceral, Pediatric and Transplantation Surgery, University Hospital RWTH Aachen, Aachen, Germany; [2]Department of Diagnostic and Interventional Radiology, University Hospital RWTH Aachen, Aachen, Germany

Correspondence to: Iakovos Amygdalos, MBBS, PhD. Department of General, Visceral, Pediatric and Transplantation Surgery, University Hospital RWTH Aachen, Pauwelsstraße 30, 52074 Aachen, Germany. Email: iamygdalos@ukaachen.de.

Comment on: Lam CSN, Bharwani AA, Chan EHY, *et al.* A machine learning model for colorectal liver metastasis post-hepatectomy prognostications. Hepatobiliary Surg Nutr 2023;12:495-506.

Keywords: Machine learning (ML); artificial intelligence (AI); colorectal; liver metastases; hepatobiliary

Submitted Apr 01, 2024. Accepted for publication Jun 19, 2024. Published online Jul 23, 2024.
doi: 10.21037/hbsn-24-187
View this article at: https://dx.doi.org/10.21037/hbsn-24-187

Colorectal cancer remains one of the commonest and most lethal malignancies worldwide (1). Up to 80% of patients with colorectal cancer develop liver metastases (CRLM), which are the main factor limiting survival (1,2). Despite advances in systemic therapy and the emergence of multimodal treatment concepts, curative surgical resection remains the best therapeutic option and provides the highest chance of long-term survival (1,2). However, liver resection is associated with perioperative morbidity and mortality, while multiple factors affect long-term prognosis (3,4). As a result, various predictive models have been developed over time, to improve preoperative patient selection and foresee unfavourable outcomes.

The first clinical risk scores (CRS) for patients undergoing resection of CRLM, were developed more than 20 years ago, with the aim to predict overall survival (OS) and recurrence-free survival (RFS) (2). Early versions by Nordlinger *et al.* (5), Fong *et al.* (6) and Nagashima *et al.* (7) comprised clinicopathological data, histological factors (such as primary tumor nodal status), size and number of liver metastases, and serum carcinoembryonic antigen (CEA) level. Later models included tumor genetic information, such as the RAS (rat sarcoma viral oncogene homolog) mutation clinical risk score (m-CS), developed by Brudvik *et al.* (8). Seeking to improve on the prognostic power and reduce bias in these models, Sasaki *et al.* (9) proposed a "Metro ticket" model, the tumor burden score (TBS). This only included the number and size of liver metastases, so that newer scores by Margonis *et al.* (10) and Chen *et al.* (11) attempted to go further, by combining TBS with histological and genetic information, resulting in the Genetic and Morphological Evaluation (GAME) and Comprehensive Evaluation of Relapse Risk (CERR) scores, respectively. Finally, Okimoto *et al.* developed the Glasgow Prognostic Score (GPS), which predicts recurrence after liver resection for CRLM, based on systemic inflammation and nutritional status (12).

In recent years, artificial intelligence (AI) and machine learning (ML) have been increasingly integrated in various areas of industry, science, and medicine (13,14). Some studies have focused on imaging modalities, such as magnetic resonance imaging (MRI), computed tomography (CT), or endoscopy, aiming to identify pathological lesions (6-8). Others have analyzed digitized histology images to predict outcomes or response to chemotherapy (13-15). Moreover, medical record digitization and development of large registries has enabled "big data" research, where various AI applications are being put to the test (13,15).

Naturally, the potential of ML algorithms to predict outcomes after liver resection for CRLM has also been

^ ORCID: 0000-0002-5138-4214.

© HepatoBiliary Surgery and Nutrition. All rights reserved.

explored in recent years. Starting in 2013, Spelt *et al.* trained an artificial neural network (ANN) on 28 pre-, intra-, and postoperative variables from a monocentric Swedish cohort of patients operated on between 1994–2009. The ANN predicted long-term survival based on six of these variables, with a C-index of 0.72 (16). In 2020, Paredes *et al.* trained two ML-models based on logistic regression with bagging to predict recurrence in a multicentric cohort of patients operated on between 2001–2018. Ten preoperative variables were selected for model training, including Kirsten rat sarcoma viral oncogene homolog (KRAS) status for one of the models. Both were found superior to the traditional Fong CRS (6) and the m-CS by Brudvik *et al.* (8), with area under the curve (AUC) ranging from 0.67 to 0.69 for 1-, 3- and 5-year recurrence (17). A different approach was used in a further study on the same cohort: here, Moro *et al.* used classification and regression trees (CART) to predict OS using 15 prognostic factors. These were ranked according to KRAS status, resulting in different combinations and two different models, both of which outperformed the Fong score. Specifically, Akaike's information criterion was better in both the KRAS-wildtype (CART 3,334 *vs.* Fong CRS 3,341) and the KRAS-mutated (CART 1,356 *vs.* Fong CRS 1,396) groups (18). Finally, in our own study on patients operated for CRLM between 2010–2021, a novel gradient-boosted decision tree (GBDT) model was trained on 24 preoperative variables to identify patients at risk of poor OS. The GBDT model ranked 6 of these variables significantly higher than the rest, so that a version trained on only these six parameters was able to identify patients at risk of significantly reduced OS (23 *vs.* 52 months, P=0.005) with a robust predictive capability (C-index 0.70) (19).

In their paper, Lam *et al.* present a prognostic ML model, based on Cox proportional hazards and least absolute shrinkage and selection operator (LASSO) regression (20). They included patients undergoing liver resection for CRLM between 2009–2018 at four centers, with data randomized 70:30 into training and validation sets and endpoints being OS and RFS. The model's predictive capability was assessed using Harrel's C-index and compared to that of the Fong CRS. From the significant risk factors identified in univariate analysis (27 for OS and 22 for RFS), 8 variables were selected for the final model: primary tumor nodal stage, neoadjuvant treatment for CRLM, Charlson Comorbidity Score, pre-hepatectomy bilirubin and CEA levels, diameter of the largest liver metastasis, extrahepatic metastasis detected on positron emission tomography (PET)-scan, and KRAS mutation status. The model, named CMAP for "CRLM Machine-learning Algorithm Prognostication", was shown to be stronger than the Fong CRS at predicting 1-year (C-index 0.651 *vs.* 0.571) and 5-year OS (C-index 0.651 *vs.* 0.574).

The study of Lam *et al.* (20) is a valuable addition to the growing field of ML-prognostication for patients undergoing resection of CRLM. The advantages of ML over traditional statistics and risk scores are discussed in the aforementioned studies and boil down to this: AI and ML models are able to analyze vast amounts of data and go through countless combinations, identifying patterns which may be missed by traditional statistics. The latter, after all, rely on the human factor (for example in the selection of variables and statistical tests), suffer from limits to the number of parameters that can be analyzed, or make assumptions about risk factors and the distribution of data.

The natural conclusion would seem to be, that traditional risk scores are outdated, and ML-based algorithms are the future. However, things are not that simple, and the current state of the art is reflected in the limitations of these studies, which are not dissimilar to those of Lam *et al.*, as acknowledged by the authors themselves (20). These include retrospective study designs, limited sample sizes, and lack of external validation. The former is accompanied by well-known issues, such as bias and missing data. The latter two are especially important for ML-algorithms, to preclude overfitting, where the algorithm is particularly well adapted to a limited dataset, but cannot make predictions on new, unknown data. Some workgroups try to overcome these problems through imputation or cross-validation, with varying results (16-19). Furthermore, studies differ in the consistency of patient cohorts and treatment strategies (e.g., eras of recruitment, frequency of major or staged resections, perioperative chemotherapy), the endpoints being studied, the way variables are included (e.g., serum CEA in continuous form *vs.* cut-off *vs.* logarithmic transformation), ML-models being employed, and statistical techniques used to assess and validate their results.

The prevalent inhomogeneity in the field is exposed by the differing variables included in models predicting the same outcomes. For example, although preoperative serum CEA, the number of CRLM, and diameter of the largest lesion persist across almost all models, perioperative chemotherapy, age, primary tumor lymph node metastases and KRAS status are only a part of some. Moreover, Lam *et al.* are the only group to include the Charlson Comorbidity Score and preoperative serum bilirubin (20), Paredes *et al.*

© HepatoBiliary Surgery and Nutrition. All rights reserved.

alone included primary tumor T-stage (17), Spelt *et al.* solely took hemorrhagic complications into account (16), whereas our model was the only one comprising body mass index and primary tumor grading (19).

These disparities highlight the need for high-quality, multicentric studies, with large and complete datasets. These would allow for comparisons between different models, external validation, as well as the prediction of outcomes other than OS/RFS, such as complications or response to chemotherapy (14,21,22). Pragmatically speaking, retrospective studies combining data from multiple centers and excluding patients with missing information are the only way to generate enough data in the short term. Concurrently, well-designed prospective studies are necessary, to maximize the quality and quantity of data made available for model training, such as genetic and epigenetic information, radiological and histological images, physiological parameters, and body composition measurements. The AI models of the future could combine all these pieces of information and take part in the decision-making process, much like a human does. The role which generative AI models (such as ChatGPT) could play in this process, is particularly intriguing.

As of the time being, the "old" is still very much useful and the "new" is not quite ready to displace it. Nevertheless, change is inevitable.

Acknowledgments

Funding: None.

Footnote

Provenance and Peer Review: This article was commissioned by the editorial office, *HepatoBiliary Surgery and Nutrition*. The article did not undergo external peer review.

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://hbsn.amegroups.com/article/view/10.21037/hbsn-24-187/coif). D.T. reports honoraria from Bayer AG. The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

References

1. Osterlund P, Salminen T, Soveri LM, et al. Repeated centralized multidisciplinary team assessment of resectability, clinical behavior, and outcomes in 1086 Finnish metastatic colorectal cancer patients (RAXO): A nationwide prospective intervention study. Lancet Reg Health Eur 2021;3:100049.
2. Joechle K, Amygdalos I, Schmidt F, et al. Value of prognostic scoring systems in the era of multimodal therapy for recurrent colorectal liver metastases. HPB (Oxford) 2023;25:1354-63.
3. Amygdalos I, Hitpass L, Schmidt F, et al. Survival after combined resection and ablation is not inferior to that after resection alone, in patients with four or more colorectal liver metastases. Langenbecks Arch Surg 2023;408:343.
4. Fernández-Moreno MC, Dorcaratto D, Garcés-Albir M, et al. Impact of type and severity of postoperative complications on long-term outcomes after colorectal liver metastases resection. J Surg Oncol 2020;122:212-25.
5. Nordlinger B, Guiguet M, Vaillant JC, et al. Surgical resection of colorectal carcinoma metastases to the liver. A prognostic scoring system to improve case selection, based on 1568 patients. Association Française de Chirurgie. Cancer 1996;77:1254-62.
6. Fong Y, Fortner J, Sun RL, et al. Clinical score for predicting recurrence after hepatic resection for metastatic colorectal cancer: analysis of 1001 consecutive cases. Ann Surg 1999;230:309-18; discussion 318-21.
7. Nagashima I, Takada T, Nagawa H, et al. Proposal of a new and simple staging system of colorectal liver metastasis. World J Gastroenterol 2006;12:6961-5.
8. Brudvik KW, Jones RP, Giuliante F, et al. RAS Mutation Clinical Risk Score to Predict Survival After Resection of Colorectal Liver Metastases. Ann Surg 2019;269:120-6.
9. Sasaki K, Morioka D, Conci S, et al. The Tumor Burden Score: A New "Metro-ticket" Prognostic Tool For

© HepatoBiliary Surgery and Nutrition. All rights reserved.

Colorectal Liver Metastases Based on Tumor Size and Number of Tumors. Ann Surg 2018;267:132-41.
10. Margonis GA, Sasaki K, Gholami S, et al. Genetic And Morphological Evaluation (GAME) score for patients with colorectal liver metastases. Br J Surg 2018;105:1210-20.
11. Chen Y, Chang W, Ren L, et al. Comprehensive Evaluation of Relapse Risk (CERR) Score for Colorectal Liver Metastases: Development and Validation. Oncologist 2020;25:e1031-41.
12. Okimoto S, Kobayashi T, Tashiro H, et al. Significance of the Glasgow Prognostic Score for patients with colorectal liver metastasis. Int J Surg 2017;42:209-14.
13. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. Lancet Oncol 2019;20:e262-73.
14. Veerankutty FH, Jayan G, Yadav MK, et al. Artificial Intelligence in hepatology, liver surgery and transplantation: Emerging applications and frontiers of research. World J Hepatol 2021;13:1977-90.
15. Amygdalos I, Hachgenei E, Burkl L, et al. Optical coherence tomography and convolutional neural networks can differentiate colorectal liver metastases from liver parenchyma ex vivo. J Cancer Res Clin Oncol 2023;149:3575-86.
16. Spelt L, Nilsson J, Andersson R, et al. Artificial neural networks--a method for prediction of survival following liver resection for colorectal cancer metastases. Eur J Surg Oncol 2013;39:648-54.
17. Paredes AZ, Hyer JM, Tsilimigras DI, et al. A Novel Machine-Learning Approach to Predict Recurrence After Resection of Colorectal Liver Metastases. Ann Surg Oncol 2020;27:5139-47.
18. Moro A, Mehta R, Tsilimigras DI, et al. Prognostic factors differ according to KRAS mutational status: A classification and regression tree model to define prognostic groups after hepatectomy for colorectal liver metastasis. Surgery 2020;168:497-503.
19. Amygdalos I, Müller-Franzes G, Bednarsch J, et al. Novel machine learning algorithm can identify patients at risk of poor overall survival following curative resection for colorectal liver metastases. J Hepatobiliary Pancreat Sci 2023;30:602-14.
20. Lam CSN, Bharwani AA, Chan EHY, et al. A machine learning model for colorectal liver metastasis post-hepatectomy prognostications. Hepatobiliary Surg Nutr 2023;12:495-506.
21. Zeng S, Li L, Hu Y, et al. Machine learning approaches for the prediction of postoperative complication risk in liver resection patients. BMC Med Inform Decis Mak 2021;21:371.
22. Lu L, Dercle L, Zhao B, et al. Deep learning for the prediction of early on-treatment response in metastatic colorectal cancer from serial medical imaging. Nat Commun 2021;12:6654.

Cite this article as: Amygdalos I, Truhn D, Vondran FWR. Outcome prediction after resection of colorectal cancer liver metastases: out with the old, in with the new? HepatoBiliary Surg Nutr 2024;13(4):732-735. doi: 10.21037/hbsn-24-187

© HepatoBiliary Surgery and Nutrition. All rights reserved.