Deep Learning in Liver Transplantation

Contributor: Rafael Calleja Lozano, César Hervás Martínez, Francisco Javier Briceño Delgado

Liver transplantation outcomes have improved. Looking for the best Donor-Recipient matching (D-R matching) as always been a challenge for the liver transplantation surgeons. Most of the proposed scores based on conventional biostatistics are not good classifiers of a problem that is considered "unbalanced." The implementation of artificial intelligence in medicine has experienced exponential growth. Deep learning, a branch of artificial intelligence with capability to handle a large number of variables with speed and multi-objective analysis. Artificial neural networks and random forests are the most widely used deep-learning classifiers in this field. Both classifiers have been able to show a high predictive ability in the graft survival of a D-R pair compared to traditional classifiers. There are even researchers that have successfully created a matching model based on one of them.

1. Current State of Art

The problem of liver donor–recipient (D–R) matching is not new and is inherent to organ transplantation. However, the Achilles' heel of liver transplantation continues to be the shortage in donors pool and the increase of waitlisted patients. A long time on waiting list (can reach as high as 20%) can lead to the death of patients waiting for an organ [1][2]. The current expansion of inclusion criteria have aggravated this problem.

The imbalance between candidates and grafts is further complicated by organ allocation policies since there exist as many policies as decisions about what to prioritize. Briefly, researchers have policies based on the principle of urgency (the "sickest-first" principle), on "individual transplant benefit principle" and "population-based transplant benefit principle" [3]. However, not all organs and not all recipients are equal. A high-risk donor-recipient matching may lead to the futility of the transplant, penalizing these patients on waiting lists. The development of perfusion machines has enabled the use of marginal grafts using normothermic machine perfusion (NMP) to serve high-risk candidates (NAPLES initiative) [4].

The classical models used to design organ allocation policies consider systems based on patient characteristics or donor risks, or a combination of donor and recipient characteristics. These D–R systems use conventional biostatistics such as logistic regression [5] but have important limitations [6]: Although these models have been analyzed in-depth, none offers an adequate response to D–R matching [3]. The reason is that these models are unable to identify the candidate on the waiting list with the highest probability of death and identify, from all available grafts, the one with the highest probability of post-transplantation success for this candidate.

In allocation policies based on the sickest-first principle, the Mayo Model for End-Stage Liver Disease (MELD) score is the most commonly used score to prioritize waitlisted candidates. Despite its utility, however, MELD (and its modifications) shows poor predictive capacity (C-statistic of 0.55) in post-transplant survival and lacks precision in prioritizing indications other than liver dysfunction (e.g., in pediatric recipients or hepatocarcinoma) [7]. This has led to the development of special systems based on extra points [8][9][10]. Other liver scoring systems, such as the Balance of Risk (BAR) score [11] or the Survival Outcome Following Liver Transplantation (SOFT) score [12], have been validated and are being used as tools in the clinical decision-making process. The BAR score is the best measure to predict 90-day morbidity with reasonable accuracy (AUC>0.70), as it can detect unfavorable D–R factor combinations before liver graft allocation [13]. Unfortunately, both BAR and SOFT are unable to identify which of several D–R pairs will achieve the best outcome; that is, they are not "matching" systems [3].

2. Concepts: Artificial Intelligence, Machine Learning and Deep Learning

AI is a branch of computational science that studies computational models capable of performing human-like activities based on two fundamental characteristics: behavior and reasoning. Its applications are diverse, including data analysis. Machine learning is defined as a branch of AI that focuses on the use of data and algorithms to mimic the way humans

learn and gradually improve the algorithms' accuracy. This learning process is understood as the ability to identify a series of complex patterns determined by a large number of variables. Therefore, the machine does not learn by itself, but the algorithm modifies itself automatically depending on the data input in its interface, thus allowing scenarios and conditions to be predicted in an automated way. For this reason, AI is being increasingly applied in the health sciences to predict clinical outcomes [14].

Machine learning and deep learning are not on the same level, but the second is part of the first. Even so, it is possible to compare both and establish some differences. While machine learning uses algorithms to analyze data, learn, and generate results or make decisions based on what it learns, deep learning structures the algorithms into layers of convolutional neural networks that help it learn and generate more accurate results. Data used by machine learning algorithms are structured and labeled for their predictions. This does not mean that they cannot work from unstructured data, but to do so, they need to perform some information pre-processing. Deep learning algorithms eliminate some of these pre-processing needs, as they can work with unstructured data and extract features in an automated or independent way. Finally, deep learning algorithms work in layers that reduce the margin of error. Each layer makes a judgment and combines that judgment with the result of the previous layer. The more information it receives and processes, the more accurate it becomes.This is the reason why AI and particularly deep-learning classifiers are an interesting alternative to traditional models [15].

3. Applications: Deep Learning in Liver Transplantation

Deep learning provides a variety of classifiers that can be utilized in almost any field of medicine [16][14][15]. Most studies on liver transplantation have focused on the development of models to predict post-transplant graft survival. ANNs and random forests are the most frequently used classifiers in this field, and studies aimed at improving D–R matching may use any of them.

3.1. Artificial Neural Networks

ANN classifiers imitate the design of human neuronal networks (**Figure 1**). Briefly, they consist of several groups of units (neurons) organized in different layers. A basic neural network consists of an input layer, a hidden layer, and an output layer. The number of layers and the ANN training can vary. Both the neurons and the relationships established between them are mathematical algorithms. The relationships (weights) between the neurons in the different layers vary as increasing data are introduced, which the model learns from. In a clinical problem, a series of input variables are introduced into the neural network, which then processes them, according to the training received, to provide output variables of clinical interest.
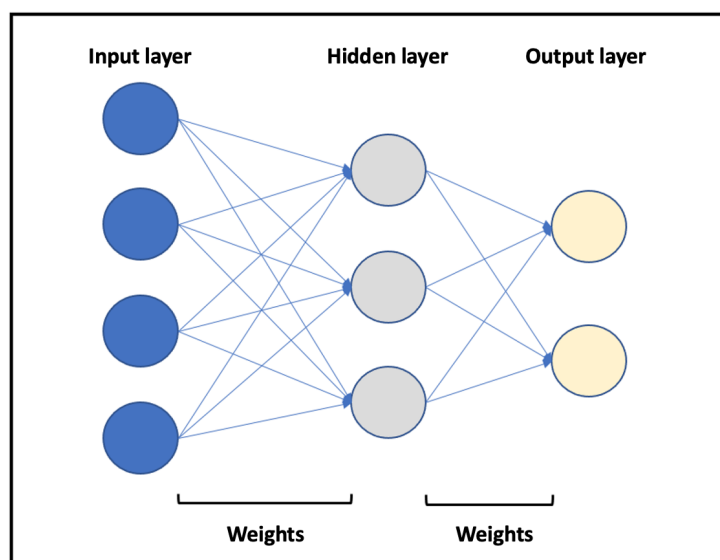
Figure 1. Representation of a basic neural network. Different layers are represented in blue (input layer), gray (hidden layer), and yellow (output layer). The arrows represent the relationships between neurons (weights).

To understand the advantages of ANNs in the field of liver transplantation, it is important to consider that the most common scenario is graft survival, while graft failure is rare, which is why it is said to be an unbalanced problem. Traditional biostatistics models are good predictors for outcomes that occur frequently; that is, they predict graft survival very well (majority class). However, they show a poor ability to predict graft failure (minority class) because it is not the usual outcome. In this regard, ANNs are able to predict both probabilities independently as they handle a large amount of

data (variables). The "surviving class or majority class" prediction is based on the concept of correct classification rate (CCR, accuracy), which refers to the proportion of training patterns classified correctly by the ANN. On the other hand, the "non-surviving class or minority class" prediction capability is measured using the concept of minimum sensitivity (MS). However is necessary to establish certain conditions for organ allocation (rules-based system). If this does not occur, the allocation would be biased, and the best candidates would receive the best grafts (i.e., those with a higher probability of success). All these concepts applied to D–R matching are schematized in **Figure 2.**
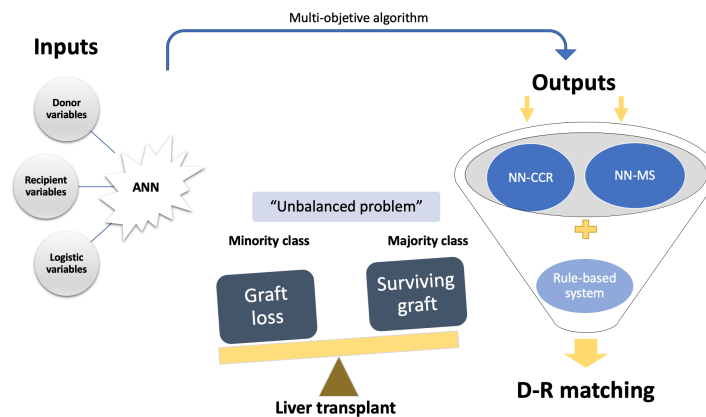
Figure 2. Diagram of an ANN-based on a multi-objective algorithm. Liver transplantation outcomes are shown as an unbalanced problem, and researchers classify them into the majority class (probability of surviving after liver transplantation, NN-CCR) and the minority class (probability of not surviving, NN-MS). By combining both probabilities (NN-CCR and NN-MS) based on input variables, researchers obtain a final D–R matching according to a rules-based system. ANN, artificial neural network; NN-CCR, neural network based on the correct classification rate or accuracy; NN-MS, neural network based on the minimum sensitivity.

From a clinical point of view, Briceño et al. [17] were the first to apply a neural network combined with a system of rules to create a donor-recipient allocation model (M.A.D.R.E model). Firstly, ANN-CCR predicted a 90.79% probability of graft survival with an area under the curve (AUC) of 0.80, while ANN-MS predicted a 71.42% probability of graft loss with an AUC of 0.82. Secondly, the authors demonstrated the superiority of ANNs in donor allocation over biostatistics-based prioritization scores (MELD, D-MELD, SOFT, P-SOFT, DRI, and BAR). Finally, the allocation system used the results obtained by the constructed ANNs and successfully assigned the best candidate for a graft according to the different probabilities (CCR and MS) from among a group of patients with higher MELD, using its rules-based system. These authors validated externally this methodology with a second study [18] achieving excellent prediction results at 3 months [CCR-AUC 0.94; MS-AUC 0.94] and 12 months (CCR-AUC 0.78; MS-AUC 0.82). The main reason for these findings was that a homogenous database with a low number of missing values was used. In their most recent study, Guijo-Rubio et al. [19] analyzed how ANNs work using the United Network for Organ Sharing (UNOS) dataset. For the 5-year endpoint, machine learning techniques, such as ANN (AUC = 0.599) or random forest (AUC = 0.644), were outperformed by logistic regression (AUC = 0.654). The predictive capacity of the AI models (including ANNs) was very similar to that obtained by traditional models (C-statistic ≤ 0.66). The reason was these classifiers were trained on a database with a high percentage of missing values.

In clinical scenarios, neural networks are very useful for finding patterns that are far too complex or numerous since they can generate near-perfect predictions using the data on which they are fit [20]. In addition, data processing is performed quickly—an essential aspect of graft allocation. However, ANNs are inherently opaque and lack interpretability because the set of weights or algorithms in hidden layers is unknown. This is called the "black-box" issue, which has made many clinicians skeptical of their use because it is necessary to know all the details of the process [16].

The predictability of an AI model depends on the robustness of the database and results are conditioned by the population in which they are trained. This is why ANNs may work very well in local and regional liver transplantation programs but cannot be extrapolated to other centers, thus requiring regional-specific ANN models. Although the results of predictive models are good, most databases are small, have a high number of missing values, or can only be applied to the population where the ANN has been trained. In addition, neural networks depend on a system of rules to properly perform donor–recipient matching but are based on the sickest-first principle. Therefore, nowadays, ANNs can only assist, but not carry out, the matching decision in all aspects of organ transplantation [21].

3.2. Random Forests

Random forests are deep-learning classifiers based on decision trees. It is an ensemble-type methodology (i.e., a model of models), and it is necessary to determine the number of models that will form the final model and verify that these models are not correlated because their results will not be very adequate. For each of the possible outputs, a different decision tree is built. The database is "split" by the researcher into different nodes. The database requires a previous treatment filter to avoid over-training and overfitting the data.

Lau et al. [22] examined how models based on random forests could predict post-transplant graft failure compared with other scores. The random forest demonstrated its superiority with an AUC of 0.787 compared to ANN (AUC = 0.734) and DRI (AUC = 0.595), as well as with other scores. The percentage of missing values in the variables used in this study, ranged from 0% to 72.22%, demonstrating the ability of random forest models to work with a high percentage of incomplete data.

The most important advantages of random forest classifiers are that a) unlike ANNs, they perform very well with a small database and a high percentage of missing values [23]; and b) they have excellent predictive power. Unfortunately, they are not useful with larger datasets (the number of decision trees they generate could be unmanageable) and they have a high risk of "over-fitting," that is, their effectiveness on the training dataset is sometimes much higher than that obtained on the validation and/or generalization dataset.

4. Current Applicability: What is on the Horizon?

AI has contributed to the field of liver transplantation through different classifiers, such as ANNs or random forests [24]. On the one hand, machine learning classifiers operate impartially as they are not affected by subjective factors. On the other, they can handle a multitude of variables of clinical interest in a quick and easy way (faster than humans) to identify the best outcome. These are the main reasons that make the use of AI so attractive from a clinical point of view. AI-based classifiers may improve D–R matching in terms such as procedure costs or individual and social benefits [25].

Deep learning is the branch of artificial intelligence that appears to be undergoing the greatest development in this field. In addition to the advances mentioned above, researchers have examples such as the application of deep learning to assess CT volumetry in living donors [26], the prediction of hepatocellular carcinoma recurrence after liver resection [27], or the identification of hepatic steatosis in living donors [28]. However, while the scientific production related to AI is more abundant in other areas of liver disease, D–R matching remains controversial. The data published to date show promising results but have not reached clinical applicability. Currently, there are three key points to implement these models in clinical decisions:

a) Overcome three ethical barriers. The first is the "black box issue," which may cause mistrust among clinicians because they do not know the weight of the variables in the models. The second is data privacy and cyber security. The last is finding an adequate answer to the following question: Who is responsible if the model fails?

b) Collect data without missing values to build large and robust datasets. However, external validation may be questionable since classifiers based on deep learning perform better in populations where they are trained. Thus, the most realistic and suitable option would be to use region-specific models.

c) Do not consider AI-based tools as "self-driving cars," but as tools to support decisions and complement current systems.

References

1. Kwong, A.; Kim, W.R.; Lake, J.R.; Smith, J.M.; Schladt, D.P.; Skeans, M.A.; Noreen, S.M.; Foutz, J.; Miller, E.; Snyder, J.J.; et al. OPTN/SRTR 2018 Annual Data Report: Liver. Am. J. Transplant. 2020, 20 (Suppl. S1), 193–299. https://doi.org/10.1111/ajt.15674.

2. Neuberger, J. Liver transplantation in the United Kingdom. Liver Transpl. 2016, 22, 1129–1135.

3. Briceno, J.; Ciria, R.; de la Mata, M. Donor-recipient matching: Myths and realities. J. Hepatol. 2013, 58, 811–820.

4. Hann, A.; Lembach, H.; Nutu, A.; Dassanayake, B.; Tillakaratne, S.; McKay, S.C.; Boteon, A.P.C.S.; Boteon, Y.L.; Mergental, H.; Murphy, N.; et al. Outcomes of normothermic machine perfusion of liver grafts in repeat liver transplantation (NAPLES ini-tiative). Br. J. Surg. 2022, 109, 372–380.

5. Lewsey, J.D.; Dawwas, M.; Copley, L.P.; Gimson, A.; Van der Meulen, J.H. Developing a prognostic model for 90-day mortal-ity after liver transplantation based on pretransplant recipient factors. Transplantation 2006, 82, 898–907.

6. Briceño, J.; Calleja, R.; Hervás, C. Artificial intelligence and liver transplantation: Looking for the best donor-recipient pair-ing. Hepatobiliary Pancreat. Dis. Int. 2022, 21, 347–353. https://doi.org/10.1016/j.hbpd.2022.03.001.

7. Doyle, H.R.; Dvorchik, I.; Mitchell, S.; Marino, I.R.; Ebert, F.H.; McMichael, J.; Fung, J.J. Predicting outcomes after liver trans-plantation. A connectionist approach. Ann. Surg. 1994, 219, 408–415. https://doi.org/10.1097/00000658-199404000-00012.

8. Schlegel, A.; Linecker, M.; Kron, P.; Gyori, G.; De Oliveira, M.L.; Mullhaupt, B.; Clavien, P.A.; Dutkowski, P. Risk Assessment in High- and Low-MELD Liver Transplantation. Am. J. Transplant. 2017, 17, 1050–1063.

9. Lai, J.C. Defining the threshold for too sick for transplant. Curr. Opin. Organ Transplant. 2016, 21, 127–132.

10. Sacleux, S.C.; Samuel, D. A Critical Review of MELD as a Reliable Tool for Transplant Prioritization. Semin. Liver Dis. 2019, 39, 403–413.

11. Dutkowski, P.; Oberkofler, C.E.; Slankamenac, K.; Puhan, M.A.; Schadde, E.; Müllhaupt, B.; Geier, A.; Clavien, P.A. Are there better guidelines for allocation in liver transplantation? A novel score targeting justice and utility in the model for end-stage liver disease era. Ann Surg. 2011, 254, 745–753. https://doi.org/10.1097/SLA.0b013e3182365081.

12. Rana, A.; Hardy, M.A.; Halazun, K.J.; Woodland, D.C.; Ratner, L.E.; Samstein, B.; Guarrera, J.V.; Brown, R.S., Jr.; Emond, J.C. Survival outcomes following liver transplantation (SOFT) score: A novel method to predict patient survival following liver transplantation. Am. J. Transplant. 2008, 8, 2537–2546. https://doi.org/10.1111/j.1600-6143.2008.02400.x

13. Boecker, J.; Czigany, Z.; Bednarsch, J.; Amygdalos, I.; Meister, F.; Santana, D.A.M.; Liu, W.J.; Strnad, P.; Neumann, U.P.; Lurje, G. Potential value and limitations of different clinical scoring systems in the assessment of short- and long-term outcome following orthotopic liver transplantation. PLoS ONE 2019, 14, e0214221.

14. Deo, R.C. Machine learning in medicine. Circulation 2015, 132, 1920–1930.

15. Fernández-Delgado, M.; Cernadas, E.; Barro, S.; Amorim, D. Do we need hundreds of classifiers to solve real world classifi-cation problems? J. Mach. Learn Res. 2014, 15, 3133–3181.

16. Veerankutty, F.H.; Jayan, G.; Yadav, M.K.; Manoj, K.S.; Yadav, A.; Nair, S.R.S.; Shabeerali, T.U.; Yeldho, V.; Sasidharan, M.; Rather, S.A. Artificial Intelligence in hepatology, liver surgery and transplantation: Emerging applications and frontiers of research. World J. Hepatol. 2021, 13, 1977–1990. https://doi.org/10.4254/wjh.v13.i12.1977.

17. Briceño, J.; Cruz-Ramírez, M.; Prieto, M.; Navasa, M.; Ortiz de Urbina, J.; Orti, R.; Gómez-Bravo, M.Á.; Otero, A.; Varo, E.; Tomé, S.; et al. Use of artificial intelligence as an innovative donor-recipient matching model for liver transplantation: Re-sults from a multicenter Spanish study. J. Hepatol. 2014, 61, 1020–1028. https://doi.org/10.1016/j.jhep.2014.05.039.

18. Ayllón, M.D., Ciria, R., Cruz-Ramírez, M., Pérez-Ortiz, M., Gómez, I., Valente, R., O'Grady, J., de la Mata, M., Hervás-Martínez, C., Heaton, N.D. and Briceño, J. (2018), Validation of artificial neural networks as a methodology for donor-recipient matching for liver transplantation. Liver Transpl, 24: 192-203. https://doi.org/10.1002/lt.24870

19. Guijo-Rubio, D.; Briceño, J.; Gutiérrez, P.A.; Ayllón, M.D.; Ciria, R.; Hervás-Martínez, C. Statistical methods versus machine learning techniques for donor-recipient matching in liver transplantation. PLoS ONE 2021, 16, e0252068. https://doi.org/10.1371/journal.pone.0252068.

20. Kelly, C.J.; Karthikesalingam, A.; Suleyman, M.; Corrado, G. King, D. Key challenges for delivering clinical impact with arti-ficial intelligence. BMC Med. 2019, 17, 195–204.

21. Ruffle, J.K.; Farmer, A.D.; Aziz, Q. Artificial intelligence-assisted gastroenterology promises and pitfalls. Am. J. Gastroenterol. 2019, 114, 422–428.

22. Lau, L.; Kankanige, Y.; Rubinstein, B.; Jones, R.; Christophi, C.; Muralidharan, V.; Bailey, J. Machine-Learning Algorithms Predict Graft Failure After Liver Transplantation. Transplantation 2017, 101, e125–e132. https://doi.org/10.1097/TP.0000000000001600.

23. Sapir-Pichhadze, R.; Kaplan, B. Seeing the Forest for the Trees: Random Forest Models for Predicting Survival in Kidney Transplant Recipients. Transplantation 2020, 104, 905–906. https://doi.org/10.1097/TP.0000000000002923.

24. Spann, A.; Yasodhara, A.; Kang, J.; Watt, K.; Wang, B.; Goldenberg, A.; Bhat, M. Applying machine learning in liver disease & transplantation: A comprehensive review. Hepatology 2020, 71, 1093–1105.

25. Sucher, R.; Sucher, E. Artificial intelligence is poised to revolutionize human liver allocation and decrease medical costs as-sociated with liver transplantation. Hepatobiliary Surg. Nutr. 2020, 9, 679–681. https://doi.org/10.21037/hbsn-20-458.

26. Park, R.; Lee, S.; Sung, Y.; Yoon, J.; Suk, H.I.; Kim, H.; Choi, S. Accuracy and Efficiency of Right-Lobe Graft Weight Estimation Using Deep-Learning-Assisted CT Volumetry for Living-Donor Liver Transplantation. Diagnostics 2022, 12, 590. https://doi.org/10.3390/diagnostics12030590.

27. Liu, Z.; Liu, Y.; Zhang, W.; Hong, Y.; Meng, J.; Wang, J.; Zheng, S.; Xu, X. Deep learning for prediction of hepatocellular carci-noma recurrence after resection or liver transplantation: A discovery and validation study. Hepatol. Int. 2022, 16, 577–589. https://doi.org/10.1007/s12072-022-10321-y.

28. Lim, J.; Han, S.; Lee, D.; Shim, J.H.; Kim, K.M.; Lim, Y.S.; Lee, H.C.; Jung, D.H.; Lee, S.G.; Kim, K.H.; et al. Identification of he-patic steatosis in living liver donors by machine learning models. Hepatol. Commun. 2022, 6, 1689–1698. https://doi.org/10.1002/hep4.1921.