Transplant International

REVIEW

Machine learning in liver transplantation: a tool for some unsolved questions?

Alberto Ferrarese[1], Giuseppe Sartori[2], Graziella Orrù[3], Anna Chiara Frigo[4], Filippo Pelizzaro[1], Patrizia Burra[1] & Marco Senzolo[1]

1 Multivisceral Transplant Unit, Department of Surgery, Oncology and Gastroenterology, Padua University Hospital, Padua, Italy
2 Forensic Neuropsychology and Forensic Neuroscience, PhD Program in Mind Brain and Computer Science, Department of General Psychology, Padua University, Padua, Italy
3 Department of Surgical, Medical, Molecular and Critical Area Pathology, University of Pisa, Pisa, Italy
4 Department of Cardiac-Thoracic-Vascular Sciences and Public Health, Biostatistics, Epidemiology and Public Health Unit, University of Padua, Padova, Veneto, Italy

Correspondence
Dr. Alberto Ferrarese MD, Multivisceral Transplant Unit, Department of Surgery, Oncology and Gastroenterology, Padua University Hospital, via Giustiniani 2, 35128 Padua, Italy.
Tel.: + 0(0)39 049.8218717;
fax: +0(0)39 049 821 7848;
e-mail:
alberto.ferrarese17@gmail.com

SUMMARY

Machine learning has recently been proposed as a useful tool in many fields of Medicine, with the aim of increasing diagnostic and prognostic accuracy. Models based on machine learning have been introduced in the setting of solid organ transplantation too, where prognosis depends on a complex, multidimensional and nonlinear relationship between variables pertaining to the donor, the recipient and the surgical procedure. In the setting of liver transplantation, machine learning models have been developed to predict pretransplant survival in patients with cirrhosis, to assess the best donor-to-recipient match during allocation processes, and to foresee postoperative complications and outcomes. This is a narrative review on the role of machine learning in the field of liver transplantation, highlighting strengths and pitfalls, and future perspectives.

Transplant International 2021; 34: 398–411

Key words
acute liver failure, cirrhosis, liver transplantation, machine learning, neural network

Received: 7 July 2020; Revision requested: 24 August 2020; Accepted: 8 January 2021; Published online:4 February 2021

The difference between inferential statistics and machine learning based models in medical research

Artificial Intelligence is an umbrella term usually indicating a field containing subsectors, such as natural language processing, deep learning, and machine learning (ML, Table 1). As most of the applications discussed here have been analyzed using ML models, this is the reason why we prefer this term throughout the manuscript.

The classical statistical approach is based on inferential analyses. Here, results are derived usually making strong assumptions about the data distribution [1]. This inferential model has been *de facto* the standard procedure for analyzing scientific experiments since 1940.

doi:10.1111/tri.13818

Table 1. Glossary.

Artificial neural network	A network of artificial neurons. Artificial neurons are built to include the basic functioning of a biological neuron. Interpretability: low
Bayesian network model	A network based on specifying relationships of conditional dependence between variables. Interpretability: high
Classification tree	Trees of decision rules with cutoffs that maximize predictive accuracy Interpretability: high
Cross-validation	Procedure that approximates the use of the predictive model on new data. The data are split randomly into a number of equally sized subsets (for example 10). The model to be evaluated is (developed and) estimated on all but one of the subsets and applied to the set that was left out. This is repeated several times, every time leaving out a different set. The error made in the prediction for the left out sets is used to evaluate the predictive ability of the model
Deep neural network	A network made of multiple levels of nonlinear operations, such as neural nets with many hidden layers Interpretability: low
Deep learning	Learning on data, using a deep neural network
Goodness of fit	Degree of overlap between actual data and predicted data, according to applied the model
Random forest	A model with multiple, randomly selected, decision trees that are averaged to make the prediction Interpretability: low
Support vector machine	ML approach used to sort two data groups, drawing lines (hyperplanes) to separate the groups according to patterns Interpretability: low

Conversely, the ML model does not make assumptions about the data structure, or on the mechanisms that generate data, making it more "agnostic" than statistical inference [2]. Further, it focuses mainly on predictive accuracy, which could be addressed also in standard statistics, but with models that are usually constrained by strong assumptions (e.g., linear regression and logistic regression). Finally, the ML approach is characterized by the widespread use of cross-validation (e.g., the procedure that approximates the use of the predictive model on new data; Table 1). Although widely used also in classical statistical models, cross-validation is a compulsory step in ML, able to reduce the risk of overfitting more than common hold-out methods (i.e., providing good results on the training set, and bad results on the validation set).

In recent years, ML has been increasingly applied in Medicine, with an exponential growth in the publications describing its use (Fig. 1). These models have consequently been widely applied in many fields of Gastroenterology and Hepatology to facilitate clinicians' diagnostic or therapeutic algorithms, or predict patient outcomes [3,4]. Examples of applications of ML in Hepatology include: predicting fibrosis in patients with viral hepatitis or nonalcoholic fatty liver disease; ascertaining the presence of esophageal varices in patients with cirrhosis; establishing the prognosis for patients with end-stage liver disease [3,5]. Certain aspects of solid organ transplantation, such as allocation, post-transplant outcome, and the management of immunosuppression, have also been explored using ML-based models [6–11].

In the last decade, there has also been interest in applying ML to liver transplantation (LT). There are two main reasons why this could theoretically be an ideal setting for ML. First, it is hard to establish the prognosis for the most common pre-LT condition, cirrhosis, because it can be influenced by several events as bacterial infection, variceal bleeding, acute kidney injury and/or hepatic encephalopathy. Second, the wide gap between donor supply and recipient demand imposes the need to optimize donor-to-recipient matching and improve postoperative graft and patient survival.

Here we provide a narrative review on the application of ML to the field of LT, highlighting strengths and pitfalls, and future perspectives (Table 2).

Predicting mortality while awaiting LT

Initially developed only to predict mortality in cirrhotic patients undergoing trans-jugular intrahepatic portosystemic shunting, the Model for End-Stage Liver Disease (MELD) score has become a reliable tool for estimating 3-month mortality for "standard" patients with end-stage liver disease [12,13]. It has consequently been introduced in many LT programs around the world as

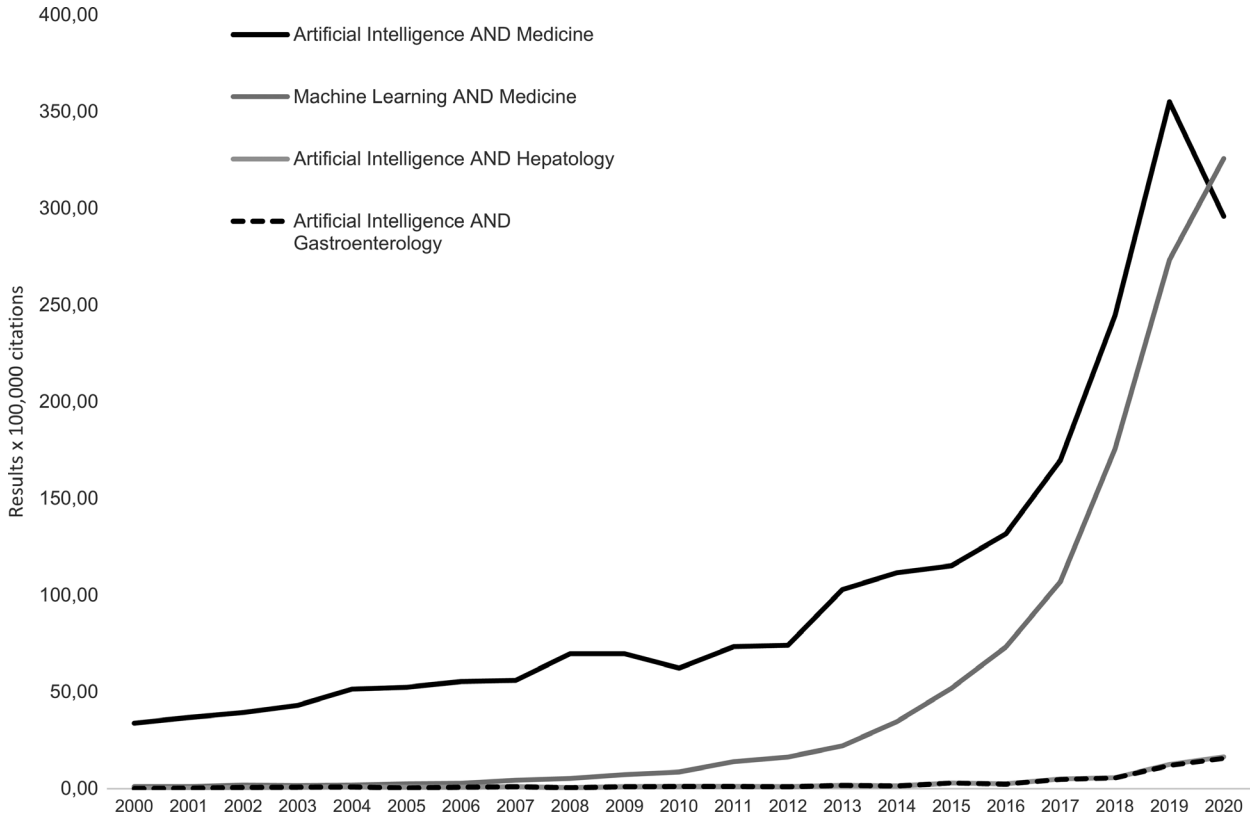

Figure 1 Trend of citations on PubMed regarding artificial intelligence and machine learning applied in medicine, gastroenterology and hepatology between 2000 and 2020.

the main tool for organ allocation, leading to a significant reduction in waiting list mortality [14]. The MELD score is less accurate, however, in capturing the prognosis for particular conditions, like refractory ascites or hepatocellular carcinoma—which commonly characterize cirrhotic patients in need of a transplant—, or for sickest candidates [15].

ML has shown promise in making the short-term prognosis for patients awaiting a transplant more reliable. Bertsimas *et al*. [16] developed a model using ML with optimal classification trees (Table 1) to predict mortality or waiting list removal in cirrhotic patients listed for LT. Using data from waitlisted patients between 2002 and 2016 in the USA, they developed an optimized prediction of mortality score, which was subsequently run in a liver simulated allocation model; it was able to predict 3-month mortality better than MELD-Na scores (AUC 0.859 vs. 0.841) or Match-MELD scores (AUC 0.859 vs. 0.823). Importantly, prediction accuracy differed between the new model and MELD-Na especially among the sickest candidates and was proven to save on average at least 418 more lives per year when compared with the currently adopted, MELD-based system.

Cucchetti *et al*. [17] further examined this topic by constructing an artificial neural network (ANN), a brain-inspired model that resembles that of biological neurons. During the learning phase, the network adjusts the weights (e.g., strengths of the synapses of the virtual neuron), increasing the predictive ability (Table 1). The ANN model was designed to predict 3-month outcome in 251 Italian patients and 137 English patients waiting for a transplant between 1999 and 2003, and included ten biochemical parameters, that are commonly recorded at time of waiting list registration. Although the two cohorts (internal and external validation sets) differed slightly regarding indications to LT or severity of underlying liver disease (their mean MELD scores were 16.7 for the Italian cohort and 14.7 for the English patients), the ANN performed better than the MELD score in predicting 3-months mortality in both groups (internal training cohort, ANN vs. MELD AUC [95% CI]: 0.98 [0.94–0.99] vs. 0.86 [0.80–0.91]; $P = 0.002$; internal validation cohort, ANN vs. MELD AUC [95% CI]: 0.95 [0.86–0.99] vs. 0.85 [0.74–0.96]; $P = 0.032$; external cohort, ANN vs. MELD AUC [95% CI]: 0.96 [0.91–0.98] vs. 0.86 [0.79–0.91]; $P = 0.04$).

Table 2. Studies on ML in the field of liver transplantation.

First author, year of publication	Study design	Aim of the study	Patients	ML models	Relevant results
Before LT					
Bertsimas, 2019 [16]	Study on the US Standard Transplant Analysis and Research Dataset (patients waitlisted between 1.2002 and 09.2016)	Prediction of 3-months patient WL mortality or removal	1 618 966 observations; 28 variables for each observation (HCC group); 25 variables (non-HCC group)	Optimal classification trees	A ML-based optimized prediction of mortality was adopted in a liver simulated allocation model, showing good predictive accuracy (AUC 0.859) External validation: no
Cucchetti, 2007 [17]	Study on Italian and English cohorts of LT candidates	Prediction of 3-months liver-disease related mortality on the WL	251 adult patients (Italy) and 137 adult patients (UK); 10 biochemical variables as input nodes	Artificial neural network	Better prediction of 3-month mortality than MELD score, both when considering the internal training cohort [AUC (95% CI): 0.98 (0.94–0.99) vs. 0.86 (0.80–0.91); $P = 0.002$]; the internal validation cohort [AUC (95% CI) 0.95 (0.86–0.99) vs. 0.85 (0.74–0.96); $P = 0.032$] and the external cohort [AUC (95% CI): 0.96 (0.91–0.98) vs. 0.86 (0.79–0.91); $P = 0.04$] External validation: yes
Speiser, 2019 [18]	Study on the US Acute Liver Failure Study Group registry (patients included 01.1998–02.2016)	Prediction of outcomes after 7 days of hospitalization in patients with ALF. Prediction of encephalopathy in patients with ALF	1042 patients with acetaminophen-induced ALF. Training dataset (n. 525 patients with 2253 observation) and test dataset (n. 517 patients with 2208 observations). Three variables collected at admission and eleven clinical daily variables	Different ML models, including binary mixed model tree	All models tested displayed good predictive accuracy. The binary mixed model tree had the best accuracy in the training dataset [0.907 (95% CI 0.894–0.918)] For the original training dataset, binary mixed model tree and support vector machine display the best AUCs (0.907 and 0.927; respectively) External validation: no.
Rajanayagam, 2013 [19]	Single-center study from Australia (pediatric patients aged 1–16 years with ALF admitted between 1991 and 2011)	Prediction of poor outcome in children with ALF admitted to a LT Center	54 pediatric patients with ALF; 18 biochemical variables as input nodes	Artificial neural network	AUCs: 1. ANN: 0.96 2. PELD-MELD score ≥42: 0.86 3. PELD-MELD score ≥27: 0.71 The Authors concluded that ANN offered a good predictive accuracy of poor outcome (death or LT) for pediatric patients with ALF. External validation: no.

Table 2. Continued.

First author, year of publication	Study design	Aim of the study	Patients	ML models	Relevant results
Allocation and short-term post-LT survival					
Briceno, 2014 [30]	Multicentric study from 11 Spanish LT Centers (LT pairs between 01.2017 and 12.2018)	Prediction of 3-months graft survival and graft loss based on D-R matching	1003 LT recipients; 57 variables (27 recipient related, 19 donor related, 12 surgery related)	Artificial neural network (a *positive survival model*, which predicts the probability of 3-months graft survival after LT; a *negative-loss model* predicts nonsurvival of the graft 3 months following LT)	Accurate prediction of graft survival (90.79%) and graft loss (71.42%) for each D-R pair. Good predictive accuracy both for positive and negative survival models (AUCs: 0.8060 and 0.8215, respectively) Balance of risk score (positive and negative survival models AUCs: 0.6799 and 0.6161, respectively) obtain the best predictive performance than commonly used scores External validation: yes (see next)
Ayllon, 2018 [31]	External validation of the model by Briceno *et al.*, in a single Center in the UK (LT pairs between 01.2002 and 12.2010)	Prediction of 3- and 12-months graft survival and graft loss based on D-R matching	858 LT recipients; 38 variables (16 recipient related, 17 donor related, 5 surgery related)	Artificial neural network (a *positive survival model*, which predicts the probability of 3-months graft survival after LT; a *negative-loss model* predicts nonsurvival of the graft three months following LT)	*Positive survival model*: prediction of 3- and 12-months post-LT graft survival AUC 0.94 and AUC 0.78, respectively *Negative-loss model*: Prediction of 3- and 12-months post-LT graft loss AUC: 0.94 and AUC 0.82, respectively The model showed a better performance than in the training database
Lau, 2018 [32]	Single-center study from Australia (LT pairs between 01.2010 and 05.2015)	Prediction of graft failure or primary non-function (e.g., death or re-LT within 30 days)	180 adult transplant patients (equally divided in training and validation group); 276 characteristics (173 donor related and 103 recipient related). Of these, the 15 most predictive were selected Training set: LT between 01.2010 and 10.2013; Validation set: LT between 11.2013 and 05.2015)	Artificial neural network; random forest	Accurate prediction of early graft failure 30-day graft failure, training set: 1. ANN: AUC 0.835 (95% CI, 0.831–0.840); 2. random forest, AUC 0.818 (95% CI, 0.812–0.824) 3-months graft failure, validation set: 1. random forests AUC 0.715 (95% CI, 0.705–0.724); 2. ANN AUC: 0.559 (95% CI, 0.548–0.569) Better performance than commonly used scores (i.e., SOFT score AUC for 30-days graft failure prediction: 0.638 [95% CI 0.632–0.645]

Table 2. Continued.

First author, year of publication	Study design	Aim of the study	Patients	ML models	Relevant results
Ershoff, 2019 [33]	Study on the US Standard Transplant Analysis and Research Dataset (LT pairs between 2005 and 2015)	3-months post-LT mortality	57 544 LT recipients (training set: 46 035; validation set 11 509). 202 variables (132 recipient related and 70 donor related)	Deep neural network	External validation: no Accuracy in predicting 3-months post-LT mortality, validation set Deep neural network AUC: [0.703 (95% CI 0.682–0.726)] BAR score AUC: 0.655 (95% CI: 0.633–0.678) SOFT score AUC: 0.688 (95% CI: 0.667–0.711)
Haydon, 2005 [34]	Single-center study from the UK, and a validation cohort (LT pairs between 1993 and 2002)	3- and 12-months post-LT patient survival	827 LT recipients from Birmingham 55 variables (37 recipient and 18 donor related). 72 input nodes	Artificial neural network (self-organizing map)	External validation: no The model was able to accurately predict high or low probabilities of survival Training set: survival probabilities at 3-months are between 0.803 [95% CI 0.754–0.847] and 0.897 [95% CI 0.843–0.937] according to donor and recipient characteristics Validation cohort: yes (2622 patients from other UK Centers, receiving LT between 1994 and 2002) Validation cohort: survival probabilities at 3-months are between 0.842 [95% CI 0.817–0.865] and 0.890 [95% CI 0.863–0.914] according to according to donor and recipient characteristics
Hoot, 2005 [35]	UNOS Registry (LT recipients between 2000 and 2002)	3-months post-LT graft survival	12 239 adult LT recipients (training set: LT between 2000 and 2001; validation set: LT in 2002), 258 variables. 30 nodes (29 nodes representing pre-LT variables, and a single dichotomous outcome node for 90-day graft survival)	Bayesian network model	Suboptimal prediction of 3-months graft survival, both in training set (AUC: 0.674) and in validation set (AUC 0.681) External validation: no

Table 2. Continued.

First author, year of publication	Study design	Aim of the study	Patients	ML models	Relevant results
Cruz-Ramirez, 2013 [54]	Multicentric study from 11 Spanish LT Centers (LT pairs between 01.2017 and 12.2018)	Creation of a multi-objective evolutionary algorithm and comparison with the mono-objective one in the setting of organ allocation	1003 LT recipients; 39 variables (20 donor related, 3 surgery related, 16 recipient related)	Artificial neural network (with multi-objective evolutionary algorithm)	In this proof-of-concept study, the Authors demonstrated a better performance of the multi-objective than the mono-objective evolutionary algorithm. Creation of a rule-based system, which can be of help in the current allocation system
Long-term survival					
Khosravi, 2015 [38]	Single-center study from Iran (LTs between 03.2008 and 03.2013)	Prediction of patient survival (1- to 5-years after LT)	1168 pediatric and adult LT recipients (n. 934 training cohort; n. 234 validation cohort)	Artificial neural network	The ANN model showed a good accuracy for 5-years post-LT patient survival [AUC (SE) 0.864 (0.043)] The Cox's proportional hazard model for 5-years post-LT survival showed an AUC [SE] equal to 0.806 [0.067] External validation: no
Dorado-Moreno, 2016 [55]	Multicentric study (7 Spanish LT Centers and 1 UK LT Center). LTs between 2007 and 2008 in Spain and between 2002 and 2010 in the UK	Prediction of post-LT (>1 year) graft survival	1406 LT pairs. 38 variables (16 recipient related, 17 donor related and 5 surgery related)	Evolutions of Artificial neural network	This proof-of-concept study showed that improvements of "classical" ANN may determine an increase in predictive accuracy on post-LT outcomes
Zhang, 2012 [56]	Single-center study from China (LTs between 02.1999 and 08.2009)	Prediction of post-LT 1-, 2-, 5-years patient mortality	290 deceased donor LT recipients with HCC; 14 variables (2 donor and 12-recipient related) and 24 input neurons	Multilayer perceptron artificial neural network	Prediction of post-LT mortality: c-statistic 1-, 2-, 5-years network: 0.909, 0.888 and 0.845, respectively External validation: no
Post-LT complications					
Hughes, 2001 [39]	Single-center study from UK (LTs between 07.1995 and 10.1997)	Prediction of post-LT (100-day) development of acute cellular rejection	124 liver grafts from 116 LTs; only biochemical variables (day post-LT, ALT level, bilirubin level, and gradients of ALT and bilirubin level)	Artificial neural network	ANN model accurately predicts acute rejection episodes 1. All rejection episodes, AUC 0.902 (95% CI 0.861–0.944) 2. Only biopsy-proven episodes, AUC 0.848 (95% CI 0.792–0.904) 3. 1-month post-LT rejection episodes, AUC 0.914 (95% CI 0.877–0.951)

Table 2. Continued.

First author, year of publication	Study design	Aim of the study	Patients	ML models	Relevant results
					4. 1 month post-LT biopsy-proven rejection episodes, AUC 0.845 (95% CI 0.785–0.905)
Lee, 2018 [45]	Single-center study from South Korea (LTs between 11.2004 and 12.2015)	Prediction of early post-LT acute kidney injury (defined as the maximal change in sCr level during first two postoperative days)	1121 recipients (367 deceased-donor LT and 844 living donor LT recipients) N. 848 in training dataset; n. 363 in validation dataset 72 variables	Different models, including neural network, random forest, support vector machine	The gradient boosting machine showed the best predictive accuracy (84%) and the best AUC (0.90; 95% CI 0.86–0.93). Most of ML models showed better accuracy than logistic regression External validation: no
Bhat, 2018 [46]	Study on the US Scientific Registry of Transplant Recipients (LTs between 10.1987 and 03.2016)	Identify predictors of post-LT new-onset diabetes within 1 year after LT	61 677 adult LT recipients	Random forest; gradient boosting, support vector machine, neural network model	High-performance random forest was the best ML model, since 88% cases were correctly classified. Age, male gender and obesity, sirolimus use were predictors of new-onset post-LT diabetes
Piscaglia, 2006 [57]	Single-center study from Spain	Prediction of significant fibrosis in hepatitis C LT recipients	510 post-LT protocol biopsies on 180 LT recipients with hepatitis C (training set n. 414; validation set n. 96); 9 clinical variables (7 of them actually used)	Artificial neural network	Good prediction of significant fibrosis [training set, AUC (95% CI) 0.87 (0.84–0.90); validation set, AUC (95% CI) 0.93 (0.86–0.97)] Logistic regression analysis-based model AUCs [training set: AUC (95% CI) 0.80 (0.76–0.84); validation set AUC (95% CI): 0.84 (0.76–0.91)] The ANN model outperformed logistic regression analysis both in training set and in validation set ($P = 0.008$ and $P = 0.045$, respectively) External validation: no

Table 2. Continued.

First author, year of publication	Study design	Aim of the study	Patients	ML models	Relevant results
Marsh, 1997 [42]	Single-center study from the USA (LTs between 1981 and 1992)	Prediction of 3-years post-LT HCC recurrence	178 patients with HCC at time of LT. 9 variables	Artificial neural network. Three different models were obtained (recurrence within 3-years; recurrence within 1-year; recurrence within 2-years)	The predictive accuracy of 1-, 2-, and 3-years recurrence varied between 0.91 and 0.95 as measured by the AUC The Authors created three groups according to probability of HCC recurrence (high probability of recurrence; high probability of nonrecurrence, intermediate group), in order to guide decision to post-LT adjuvant chemotherapy External validation: no
Marsh, 2003 [43]	Single-center study from the USA (LTs between 1991 and 1994)	Prediction of post-LT HCC recurrence	103 patients with HCC in explanted liver	Artificial neural network (previously developed by the same group, see above) and tissue genotyping	Combined application of ANN and tissue genotyping gave predictions of 91/103 patients, increasing the discriminatory power of ANN by 15% The accuracy was 100% in the 81 out of 91 patients who were alive and regularly followed-up External Validation: yes (see next)
Rodriguez-Luna, 2005 [44]	Single-center study from the USA (LTs between 1999 and 2002)	Validation of ML + tissue genotyping model for prediction of post-LT HCC recurrence	19 patients with HCC in explanted liver	Artificial neural network and tissue genotyping	Prediction of 3-years recurrence equal to 100% (3 patients) Combination of ANN and tissue genotyping allowed prediction in 17/19 patients (discriminatory power of the combined analysis equal to 89.5%)

ALF, acute liver failure; ANN, artificial neural network; AUC, area under the curve; HCC, hepatocellular carcinoma; LT, liver transplantation; MELD, model for end-stage liver disease; WL, waiting list.

ML-based models have been used to predict survival in noncirrhotic patients awaiting a transplant as well. In a proof-of-concept study, Speiser *et al.* [18] showed that ML could adequately predict day-by-day outcome in patients with acetaminophen-induced acute liver failure (ALF). The Authors included 1042 patients with ALF between 1998 and 2016, and divided them in a training and an internal dataset. Identifying as the primary endpoint the occurrence of encephalopathy within the first 7 days of hospitalization, the study showed that extensions or variants of random forest (Table 1) accurately predicted patient outcomes. The best accuracy was provided by the binary mixed model tree in the training dataset [AUC: 0.907 (95% CI 0.894–0.918)], and by binary mixed model tree and support vector machine (Table 1) in the validation dataset (AUCs equal to 0.907 and 0.927; respectively), even if an external validation was not provided.

An ANN-based model was applied to 54 pediatric patients with ALF, too, showing a good prognostic accuracy when compared with the commonly used scores (ANN vs. MELD-PELD peak score ≥42 AUC: 0.96 vs. 0.86), but the small sample size and the wide study period (over 10 years) prevented any robust conclusions from being drawn [19].

In summary, few studies are currently available on the role of ML in predicting outcomes in candidates for LT. Some interesting results have emerged for patients with cirrhosis, with preliminary models showing a similar—or perhaps better—predictive accuracy than the commonly used scores. This would theoretically be of great interest for the purposes of organ allocation, but deserve further exploration, especially among patients with ALF and pediatric populations. Moreover, the absence of external validation for many of available studies may represent another pitfall in the interpretation of currently published results.

Optimizing organ allocation and improving post-LT short-term survival

The rationale behind organ allocation systems is to maximize the use of available organs and reduce the mortality of patients on the waiting list [20,21]. Theoretically, organ allocation may be driven by three important principles: urgency (allocation to the patient estimated to have the shortest survival without a transplant); utility (allocation to the patient estimated to have the longest post-transplant survival); or transplant benefit (allocation based on the difference between the mean survival estimates with and without a transplant). Despite several efforts, a unified and standardized international model has yet to be adopted. This is partly due to significant ethical and socio-cultural differences around the world [22,23], and also because only about 25% of patients undergo "benchmark" transplantations, with a wide variability across centers [24,25]. Therefore, a model that can perform well for one population may be less appropriate for others.

Most LT programs adopting an urgency-based allocation algorithm rely on a purely biochemical system based on the MELD score, or subsequent revisions [14,26]. In utility-based organ allocation systems, the MELD score is not a reliable tool because it is a weak predictor of post-LT mortality. In this setting, where appropriate donor–recipient pairing is important in order to improve outcomes, several scores, which consider the characteristics of donors [27], or both donors and recipients [28,29] have been developed. Even if such scores derive from objective factors that are readily available at time of organ allocation, whether post-transplant outcome can be predicted from just a handful of variables remains debatable.

Briceño *et al.* [30] applied ML to the complex scenario of organ allocation, combining 57 donor-, recipient-, or surgery-related variables in an ANN model. Using the outcomes of 1003 patients who received a graft in Spain between 2007 and 2008, they showed that the ANN model was more accurate than commonly used logistic regression models in predicting 3-month graft survival (ANN vs. MELD AUC: 0.80 vs. 0.50, $P = 0.001$) and graft loss (ANN vs. MELD AUC: 0.82 vs. 0.41; $P = 0.001$) for each donor-to-recipient pair. This better accuracy was confirmed also after comparing the ML model with above-mentioned scores [28,29] (graft survival: ANN vs. balance of risk AUC 0.80 vs. 0.67; $P = 0.001$; graft loss: ANN vs. balance of risk AUC: 0.82 vs. 61; $P = 0.001$).

This model was further validated in an English cohort of 858 patients who underwent LT between 2002 and 2010 (patients with HCC were ruled out) [31]. Its accuracy in predicting 3- and 12-month graft survival (ANN AUC: 0.94 and 0.78, respectively) and graft loss (ANN AUC: 0.94 and 0.82, respectively) was significantly better than that of the commonly used scores, being balance of risk the second-best score (3- and 12-month graft survival AUC: 0.84 and 0.71, respectively). The model also seemed to perform better than in the training cohort, probably due to differences between the two cohorts' baseline characteristics.

Another experience in this field came from Australia, where two ML-based models were developed to predict

early graft survival, considering the characteristics of 180 deceased donor LT recipients transplanted between 2010 and 2015 [32]. The models included 15 of 276 baseline donor and recipient variables, and exhibited a good accuracy in predicting graft failure at 1 month (random forest AUC [95% CI] 0.818 [0.812–0.824]; ANN AUC [95% CI]: 0.835 [0.831–0.840], respectively), in a better way than other scores obtained with logistic regression analyses (Donor risk index [27] AUC [95% CI] 0.680 [0.669–0.690]; SOFT score AUC [95% CI]: 0.638 [0.632–0.645], respectively).

Several studies attempted to apply ML to the prediction of graft and/or patient survival at 3 months, obtaining suboptimal results in terms of accuracy [33–35]. This might be because donor variables could influence the early post-transplant phase more than in subsequent months (when complications unrelated to the liver might also occur). Although these studies demonstrated that ML-based models were more accurate than commonly adopted scores, the gain was often not clinically relevant. In the study by Lau *et al.* [32], for instance, the model previously applied for prediction of 30-day graft failure provided a less accurate prediction when the endpoint was the 90-day outcome [random forests and ANN AUCs (92% CI): 0.715 (0.705–0.724) and 0.559 (0.548–0.569), respectively].

In summary, ML will presumably be a useful tool for improving organ allocation and predicting short-term graft and patient survival in the next future. Nevertheless, the available studies provided information on patients transplanted over a broad period of time, with different liver disease etiologies and involving different donors from those being managed in the future. Therefore, the applicability of ML-based prognostics to current or future LT cohorts remains to be seen.

Predicting post-LT long-term outcome and post-LT complications

Long-term outcome after solid organ transplantation is even more difficult to predict than in the early post-transplant period because it may also be influenced by conditions unrelated to the graft, such as infections, malignancies and metabolic or cardiovascular diseases [36,37]. Khosravi *et al.* [38] used an ANN-based model to predict long-term outcome (beyond one year after LT) in 1168 patients >2 years old who underwent LT (10.7% were cases of living donor LT) between 2008 and 2013. The most accurate model included 16 of 37 baseline predictors (five recipient characteristics, 10 intra or postoperative variables, and chronic rejection), and afforded an accurate prognosis on patient survival between 1 and 5 years after LT [AUC (standard error): 0.864 (0.043)], similar to what was retrieved by the "standard" Cox's proportional hazards model [AUC (standard error) 0.806 (0.067)].

More results are available on the use of ML to foresee commonly encountered complications after LT. Hughes *et al.* [39] applied an ANN model to predict acute rejection early after LT in 117 adult recipients, using biochemical characteristics (ALT, bilirubin and their dynamic course) and timing since LT. The model showed a good accuracy [AUC (95% CI) 0.902 (0.861–0.944)], even if results might be interpreted with caution, since seven episodes of acute rejection were not biopsy-proven, the liver disease etiologies were heterogeneous, and there were some cases of re-LT. Nevertheless, this was an important proof-of-concept study for larger future investigations.

Hepatocellular carcinoma (HCC) recurrence has been demonstrated in 6–18% of patients after LT, with a significant impact on patient survival [40]. Composite models of HCC recurrence, considering morphological, clinical and biochemical characteristics, have been proposed with a view to optimizing post-LT surveillance, stratifying patients' risk and tailoring their immunosuppressant therapy [41]. Marsh *et al.* [42] developed an ANN model for predicting HCC recurrence at 1, 2, and 3 years post-LT, based on five risk factors (sex, tumor number, size and intrahepatic distribution, and grade of vascular invasion) retrieved from 178 LT recipients. The ML model allowed to stratify patients into three groups with different risk of recurrence, in order to theoretically deserve a tailored postoperative surveillance. The same group adopted the previously developed ANN model in combination with tissue genotyping for microsatellite mutations or deletions in 103 explanted livers with HCC [43]. This combination increased by 15% the predictive accuracy of ANN regarding post-LT HCC recurrence (88% vs. 71%). The results were externally validated by Rodriguez-Luna *et al.* [44] in a small cohort of patients transplanted between 1992 and 2002 in another American LT Center. The Author confirmed that this composite model correctly predicted post-LT HCC recurrence in 17/19 (89.5%) patients.

Some innovative results have come from the application of ML regarding metabolic or renal complications after LT. A study by Lee *et al.* [45] investigated the probability of postoperative acute kidney injury (defined as a maximal change in serum creatinine level during the first two days after surgery). According to their findings, ML-based model performed better than the

Transplant International 2021; 34: 398–411

standard statistical logistic regression model [random forest vs. logistic regression analysis: AUC (95% CI): 0.61 (0.56–0.66) vs. 0.85 (0.81–0.89); $P < 0.001$], and showed that cold ischemia time and intraoperative mixed venous oxygen saturation were the most important variables associated with renal dysfunction. The inclusion of both living and deceased donor transplants, and the small number of cases considered should be taken into account when interpreting their findings, however.

A further study conducted at by Bhat *et al.* [46] explored the probability of new-onset diabetes after LT in a large cohort of patients coming from the US transplant registry. The Authors demonstrated that a high-performance random forest model was able to predict diabetes in 88% patients within a year after their transplant. The risk of new-onset diabetes rose by 33% when sirolimus was used instead of tacrolimus. The model also showed that diabetes carried a 55% higher risk of death at 10 years.

In summary, ML is attracting attention as an innovative tool for predicting long-term post-LT complications. ML models may pave the way to a personalized post-transplant follow-up, taking individual pre- and post-LT features into account. Many of the above-mentioned studies are proofs of concept, so their preliminary results need to be further explored and confirmed. More research is needed in the prediction of long-term follow-up after transplantation.

Conclusions and future perspectives

As in other fields of Medicine and Gastroenterology and Hepatology, ML will probably influence the clinical setting of LT in the near future. ML modeling of LT datasets could improve prognostic accuracy, and the applicability of the model's predictions to new cases. It could also facilitate the selection of the most influential predictors from among the numerous variables commonly collected from donors and recipients. This would aid clinicians in many settings, both before and after surgery, improving patients' outcomes and quality of life, and fostering a personalized follow-up.

Several topics pertaining to the field of decompensated cirrhosis will probably be further explored using ML, such as the clinical course of acute-on-chronic liver failure, or ALF, bacterial infections, recurrence of variceal bleeding or hepatic encephalopathy. This will be helpful when it comes to considering a patient for a transplant. We can also expect a further refinement of prognostics at the time of surgery to be achieved by exploring the use of extended-criteria donors or grafts retrieved from donors after circulatory death. Moreover, it would be useful to be able to predict disease recurrence (in cases of cholestatic and autoimmune disease), tumor recurrence, or *de novo* cancers, also to improve the management of immunosuppression (as already seen in kidney transplant recipients) and patient adherence [47–51] (Table 3).

Table 3. Potential fields of interest for future development of ML before and after liver transplantation.

Before LT
Course of acute-on-chronic liver failure
Resolution of bacterial infections
During or soon after surgery
Allocation of extended-criteria donors and donors after circulatory death
After LT
Primary disease recurrence (e.g., cholestatic diseases)
Immunosuppression management
Adherence after LT in special populations
De novo tumors after liver transplantation

The data produced by ML models should nonetheless be interpreted with caution. Some models are unable to shed light on the real contribution of a given factor, or reveal how changing a given variable will affect the model. It is also still unclear whether the accuracy of some models is reproducible in cohorts with different characteristics. It is worth noting that the number and type of donor and/or recipient variables used across studies, as well as the types of ML model, differ significantly, so the results of these studies are not comparable with one another. Finally, one *caveat* of ML modeling lies in the difficulty of interpreting the output of some of the models used to analyze transplantation datasets (such as ANNs and random forests). Other already available models derived from regression analyses, as MELD score formula in organ allocation, may still offer more readily interpretable decision-making rules that can be incorporated in the clinical decision-making process.

Finally, although ML can make accurate predictions, it is ultimately up to health care workers to make decisions based on their patient's characteristics, clinical condition, and expectations [3,52]. This is particularly important in the setting of organ transplantation, where issues of ethics and justice are of the utmost importance and cannot be categorized as mere computational variables. Transplantation is not just about installing spare parts, like a puzzle, and the human mind will retain a leading role in this process [53].

Funding

The authors have declared no funding.

Conflict of interest

The authors have declared no conflicts of interest.

REFERENCES

1. Breiman L. Random forests. *Mach Learn* 2001; **45**: 5.
2. Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. *Nat Methods* 2018; **15**: 233.
3. Le Berre C, Sandborn WJ, Aridhi S, *et al*. Application of artificial intelligence to gastroenterology and hepatology. *Gastroenterology* 2020; **158**: 76.
4. Ruffle JK, Farmer AD, Aziz Q. Artificial intelligence-assisted gastroenterology – promises and pitfalls. *Am J Gastroenterol* 2019; **114**: 422.
5. Spann A, Yasodhara A, Kang J, *et al*. Applying machine learning in liver disease and transplantation: a comprehensive review. *Hepatology* 2020; **71**: 1093.
6. Niel O, Bastard P. Artificial intelligence improves estimation of tacrolimus area under the concentration over time curve in renal transplant recipients. *Transpl Int* 2018; **31**: 940.
7. Thishya K, Vattam KK, Naushad SM, Raju SB, Kutala VK. Artificial neural network model for predicting the bioavailability of tacrolimus in patients with renal transplantation. *PLoS One* 2018; **13**: e0191921.
8. Sheppard D, McPhee D, Darke C, *et al*. Predicting cytomegalovirus disease after renal transplantation: an artificial neural network approach. *Int J Med Inform* 1999; **54**: 55.
9. Llorca J, Dierssen-Sotos T, Gómez-Acebo I, Gonzalez-Castro A, Miñambres E. Artificial neural networks predict mortality after lung transplantation better than logistic regression. *J Heart Lung Transplant* 2009; **28**: 1237.
10. Hearn J, Ross HJ, Mueller B, *et al*. Neural networks for prognostication of patients with heart failure. *Circ Heart Fail* 2018; **11**: e005193.
11. Yoon J, Zame WR, Banerjee A, Cadeiras M, Alaa AM, van der Schaar M. Personalized survival predictions via trees of predictors: an application to cardiac transplantation. *PLoS One* 2018; **13**: e0194985.
12. Kamath PS, Kim WR, Group ALDS. The model for end-stage liver disease (MELD). *Hepatology* 2007; **45**: 797.
13. Said A, Williams J, Holden J, *et al*. Model for end stage liver disease score predicts mortality across a broad spectrum of liver disease. *J Hepatol* 2004; **40**: 897.
14. Kamath PS, Wiesner RH, Malinchoc M, *et al*. A model to predict survival in patients with end-stage liver disease. *Hepatology* 2001; **33**: 464.
15. Biggins SW, Bambha K. MELD-based liver allocation: who is underserved? *Semin Liver Dis* 2006; **26**: 211.
16. Bertsimas D, Kung J, Trichakis N, Wang Y, Hirose R, Vagefi PA. Development and validation of an optimized prediction of mortality for candidates awaiting liver transplantation. *Am J Transplant* 2019; **19**: 1109.
17. Cucchetti A, Vivarelli M, Heaton ND, *et al*. Artificial neural network is superior to MELD in predicting mortality of patients with end-stage liver disease. *Gut* 2007; **56**: 253.
18. Speiser JL, Karvellas CJ, Wolf BJ, Chung D, Koch DG, Durkalski VL. Predicting daily outcomes in acetaminophen-induced acute liver failure patients with machine learning techniques. *Comput Methods Programs Biomed* 2019; **175**: 111.
19. Rajanayagam J, Frank E, Shepherd RW, Lewindon PJ. Artificial neural network is highly predictive of outcome in paediatric acute liver failure. *Pediatr Transplant* 2013; **17**: 535.
20. Freeman RB, Jamieson N, Schaubel DE, Porte RJ, Villamil FG. Who should get a liver graft? *J Hepatol* 2009; **50**: 664.
21. Neuberger J. The introduction of MELD-based organ allocation impacts 3-month survival after liver transplantation by influencing pretransplant patient characteristics. *Transpl Int* 2009; **22**: 979.
22. Dutkowski P, Clavien PA. Scorecard and insights from approaches to liver allocation around the world. *Liver Transpl* 2016; **22**: 9.
23. Tschuor C, Ferrarese A, Kuemmerli C, *et al*. Allocation of liver grafts worldwide – is there a best system? *J Hepatol* 2019; **71**: 707.
24. Muller X, Marcon F, Sapisochin G, *et al*. Defining benchmarks in liver transplantation: a multicenter outcome analysis determining best achievable results. *Ann Surg* 2018; **267**: 419.
25. Carbone M, Nardi A, Marianelli T, *et al*. International comparison of liver transplant programmes: differences in indications, donor and recipient selection and outcome between Italy and UK. *Liver Int* 2016; **36**: 1481.
26. Biggins SW, Kim WR, Terrault NA, *et al*. Evidence-based incorporation of serum sodium concentration into MELD. *Gastroenterology* 2006; **130**: 1652.
27. Feng S, Goodrich NP, Bragg-Gresham JL, *et al*. Characteristics associated with liver graft failure: the concept of a donor risk index. *Am J Transplant* 2006; **6**: 783.
28. Rana A, Hardy MA, Halazun KJ, *et al*. Survival outcomes following liver transplantation (SOFT) score: a novel method to predict patient survival following liver transplantation. *Am J Transplant* 2008; **8**: 2537.
29. Dutkowski P, Oberkofler CE, Slankamenac K, *et al*. 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; discussion 53.
30. Briceño J, Cruz-Ramírez M, Prieto M, *et al*. Use of artificial intelligence as an innovative donor-recipient matching model for liver transplantation: results from a multicenter Spanish study. *J Hepatol* 2014; **61**: 1020.
31. Ayllón MD, Ciria R, Cruz-Ramírez M, *et al*. Validation of artificial neural networks as a methodology for donor-recipient matching for liver transplantation. *Liver Transpl* 2018; **24**: 192.
32. Lau L, Kankanige Y, Rubinstein B, *et al*. Machine-learning algorithms predict graft failure after liver transplantation. *Transplantation* 2017; **101**: e125.
33. Ershoff BD, Lee CK, Wray CL, *et al*. Training and validation of deep neural networks for the prediction of 90-day post-liver transplant mortality using UNOS registry data. *Transplant Proc* 2020; **52**: 246.

34. Haydon GH, Hiltunen Y, Lucey MR, *et al*. Self-organizing maps can determine outcome and match recipients and donors at orthotopic liver transplantation. *Transplantation* 2005; **79**: 213.
35. Hoot N, Aronsky D. Using Bayesian networks to predict survival of liver transplant patients. *AMIA Annu Symp Proc* 2005; **2005**: 345.
36. Durand F. How to improve long-term outcome after liver transplantation? *Liver Int* 2018; **38**(Suppl 1): 134.
37. Watt KD, Pedersen RA, Kremers WK, Heimbach JK, Charlton MR. Evolution of causes and risk factors for mortality post-liver transplant: results of the NIDDK long-term follow-up study. *Am J Transplant* 2010; **10**: 1420.
38. Khosravi B, Pourahmad S, Bahreini A, Nikeghbalian S, Mehrdad G. Five years survival of patients after liver transplantation and its effective factors by neural network and cox poroportional hazard regression models. *Hepat Mon* 2015; **15**: e25164.
39. Hughes VF, Melvin DG, Niranjan M, Alexander GA, Trull AK. Clinical validation of an artificial neural network trained to identify acute allograft rejection in liver transplant recipients. *Liver Transpl* 2001; **7**: 496.
40. Welker MW, Bechstein WO, Zeuzem S, Trojan J. Recurrent hepatocellular carcinoma after liver transplantation – an emerging clinical challenge. *Transpl Int* 2013; **26**: 109.
41. Berenguer M, Burra P, Ghobrial M, *et al*. Posttransplant management of recipients undergoing liver transplantation for hepatocellular carcinoma. Working Group Report from the ILTS Transplant Oncology Consensus Conference. *Transplantation* 2020; **104**: 1143.
42. Marsh JW, Dvorchik I, Subotin M, *et al*. The prediction of risk of recurrence and time to recurrence of hepatocellular carcinoma after orthotopic liver transplantation: a pilot study. *Hepatology* 1997; **26**: 444.
43. Marsh JW, Finkelstein SD, Demetris AJ, *et al*. Genotyping of hepatocellular carcinoma in liver transplant recipients adds predictive power for determining recurrence-free survival. *Liver Transpl* 2003; **9**: 664.
44. Rodriguez-Luna H, Vargas HE, Byrne T, Rakela J. Artificial neural network and tissue genotyping of hepatocellular carcinoma in liver-transplant recipients: prediction of recurrence. *Transplantation* 2005; **79**: 1737.
45. Lee HC, Yoon SB, Yang SM, *et al*. Prediction of acute kidney injury after liver transplantation: machine learning approaches vs. logistic regression model. *J Clin Med* 2018; **7**: 428
46. Bhat V, Tazari M, Watt KD, Bhat M. New-onset diabetes and preexisting diabetes are associated with comparable reduction in long-term survival after liver transplant: a machine learning approach. *Mayo Clin Proc* 2018; **93**: 1794.
47. Seymour CW, Kennedy JN, Wang S, *et al*. Derivation, validation, and potential treatment implications of novel clinical phenotypes for sepsis. *JAMA* 2019; **321**: 2003.
48. Zheng MH, Shi KQ, Lin XF, *et al*. A model to predict 3-month mortality risk of acute-on-chronic hepatitis B liver failure using artificial neural network. *J Viral Hepat* 2013; **20**: 248.
49. Tang J, Liu R, Zhang YL, *et al*. Corrigendum: application of machine-learning models to predict tacrolimus stable dose in renal transplant recipients. *Sci Rep* 2018; **8**: 46936.
50. Connor JP, Symons M, Feeney GF, Young RM, Wiles J. The application of machine learning techniques as an adjunct to clinical decision making in alcohol dependence treatment. *Subst Use Misuse* 2007; **42**: 2193.
51. Lo-Ciganic WH, Donohue JM, Thorpe JM, *et al*. Using machine learning to examine medication adherence thresholds and risk of hospitalization. *Med Care* 2015; **53**: 720.
52. Rampton V. Artificial intelligence versus clinicians. *BMJ* 2020; **369**: m1326.
53. Starzl TE. *The Puzzle People Memoirs of a Transplant Surgeon*, Pittsburgh, USA: University of Pittsburgh Press, 1992.
54. Cruz-Ramírez M, Hervás-Martínez C, Fernández JC, Briceño J, de la Mata M. Predicting patient survival after liver transplantation using evolutionary multi-objective artificial neural networks. *Artif Intell Med*. 2013; **58**: 37–49.
55. Dorado-Moreno M, Pérez-Ortiz M, Gutiérrez PA, Ciria R, Briceño J, Hervás-Martínez C. Dynamically weighted evolutionary ordinal neural network for solving an imbalanced liver transplantation problem. *Artif Intell Med*. 2017; **77**: 1–11.
56. Zhang M, Yin F, Chen B, et al. Mortality risk after liver transplantation in hepatocellular carcinoma recipients: a nonlinear predictive model. *Surgery*. 2012; **151**: 889–97.
57. Piscaglia F, Cucchetti A, Benlloch S, et al. Prediction of significant fibrosis in hepatitis C virus infected liver transplant recipients by artificial neural network analysis of clinical factors. *Eur J Gastroenterol Hepatol*. 2006; **18**: 1255–61.