

Review

Open Access



# From prediction to practice: a narrative review of recent artificial intelligence applications in liver transplantation

Khush Patel<sup>1,2</sup>, Ashton A. Connor<sup>1,2,3</sup>, Sudha Kodali<sup>4,5,6,7</sup>, Constance M. Mobley<sup>1,2,3</sup>, David Victor<sup>4,5,6,7</sup>, Mark J. Hobeika<sup>1,2,3</sup>, Youssef Dib<sup>1,2</sup>, Ashish Saharia<sup>1,2,3</sup>, Yee Lee Cheah<sup>1,2,3</sup>, Caroline J. Simon<sup>1,2,3</sup>, Elizabeth W. Brombosz<sup>1,2</sup>, Linda W. Moore<sup>1,2,3</sup>, R. Mark Ghobrial<sup>1,2,3,5,7</sup>

<sup>1</sup>Department of Surgery, Houston Methodist Hospital, Houston, TX 77030, USA.

<sup>2</sup>J.C. Walter Jr. Transplant Center, Houston Methodist Hospital, Houston, TX 77030, USA.

<sup>3</sup>Department of Surgery, Weill Cornell Medical College, New York, NY 10065, USA.

<sup>4</sup>Department of Medicine, Houston Methodist Hospital, Houston, TX 77030, USA.

<sup>5</sup>Department of Medicine, Weill Cornell Medical College, New York, NY 10065, USA.

<sup>6</sup>Sherrie and Alan Conover Center for Liver Disease and Transplantation, Houston, TX 77030, USA.

<sup>7</sup>Underwood Center for Digestive Health, Houston Methodist Hospital, Houston, TX 77030, USA.

**Correspondence to:** Dr. R. Mark Ghobrial, J.C. Walter Jr. Transplant Center, Houston Methodist Hospital, 6550 Fannin St, SM1661, Houston, TX 77030, USA. E-mail: RMGhobrial@houstonmethodist.org

**How to cite this article:** Patel K, Connor AA, Kodali S, Mobley CM, Victor D, Hobeika MJ, Dib Y, Saharia A, Cheah YL, Simon CJ, Brombosz EW, Moore LW, Ghobrial RM. From prediction to practice: a narrative review of recent artificial intelligence applications in liver transplantation. *Art Int Surg*. 2025;5:298-321. <https://dx.doi.org/10.20517/ais.2024.103>

**Received:** 7 Dec 2024 **First Decision:** 21 Mar 2025 **Revised:** 6 May 2025 **Accepted:** 12 May 2025 **Published:** 20 Jun 2025

**Academic Editors:** Andrew Gumbs, Mohammad Abu-Hilal **Copy Editor:** Pei-Yun Wang **Production Editor:** Pei-Yun Wang

## Abstract

Liver transplantation (LT) is the definitive treatment for end-stage liver disease and certain liver cancers. This involves complex decision making across the transplant continuum. Artificial intelligence (AI), with its ability to analyze high-dimensional data and derive meaningful patterns, shows promise as a transformative tool to address these challenges. In this narrative review, we searched PubMed from January 2021 to October 2024 using keywords such as “artificial intelligence”, “machine learning”, “deep learning”, and “liver transplantation”. Only full-text, English-language studies on adult populations (with minimum sample sizes deemed appropriate by each study’s design) were included, with a total of 65 articles. These publications examined AI applications in pre-transplant risk assessment (9), donor liver assessment (11), transplant oncology (11), graft survival prediction (7), overall survival prediction (11), immunosuppression management (4), and post-transplant risk prediction (12). Tree-based methods showed high accuracy in predictive tasks, while deep learning excelled in medical imaging analysis. Despite these advancements, only 6% of studies addressed algorithmic fairness, and 41% of neural



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



network implementations lacked interpretability methods. Key challenges included data harmonization, multicenter validation, and integration with existing clinical workflows. Despite these limitations, AI continues to show promise for optimizing critical steps along the LT continuum. As the field progresses, the focus must remain on using AI to expand access and optimize care, ensuring it supports rather than restricts transplant opportunities.

**Keywords:** Artificial intelligence, machine learning, liver transplantation, clinical decision support

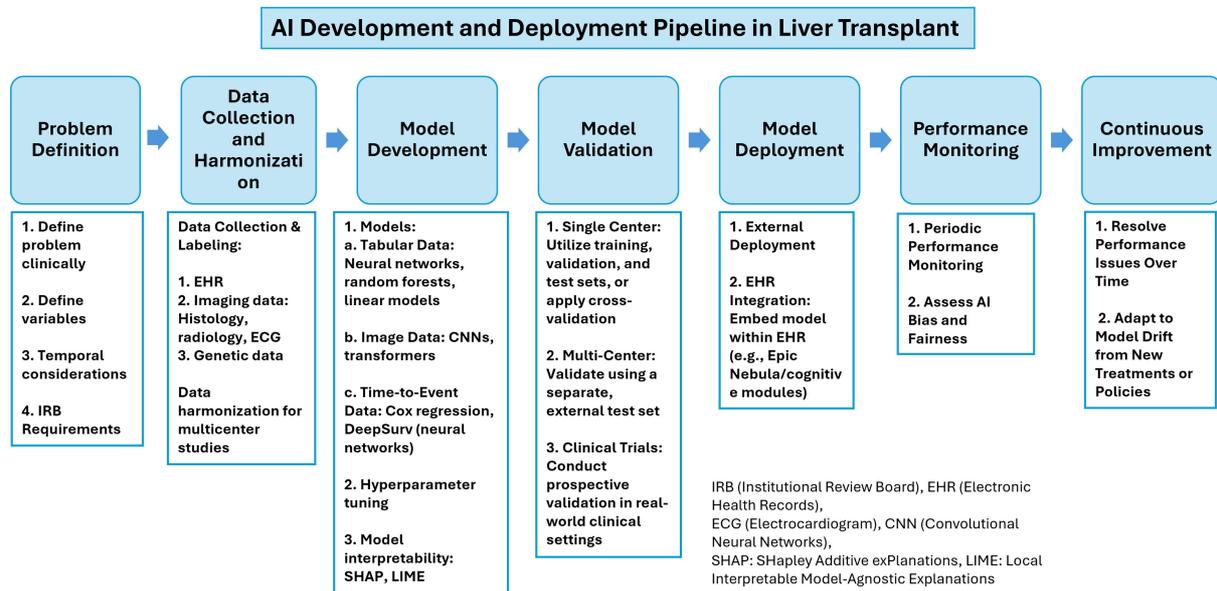
## INTRODUCTION

Liver transplantation (LT) is a life-saving procedure for patients with end-stage liver disease and specific liver cancers, often serving as the only curative option. Its success relies on navigating complex decisions, from candidate selection and donor assessment to post-transplant care. Clinicians must process vast, heterogeneous data, yet traditional tools often fail to capture intricate relationships or deliver personalized predictions. Artificial intelligence (AI) has emerged as a promising solution to address these challenges in LT by analyzing complex patterns across diverse data types, including clinical variables, imaging, and molecular markers<sup>[1-3]</sup>. At its core, AI enhances decision-making accuracy and patient outcomes by leveraging historical data to predict new cases.

The successful implementation of AI in LT involves seven key phases [Figure 1]: problem definition, data collection, model development, validation, deployment, performance monitoring, and continuous improvement. This process begins with defining the clinical question and identifying data sources like electronic health records (EHRs), imaging studies, and lab results. During model development, algorithms are trained on historical data to detect patterns predicting outcomes. Hyperparameter optimization is the process of selecting the best combination of parameters that control the learning process of a model to maximize its performance. Validation ensures the model generalizes across centers. Deployment integrates AI into clinical workflows via user-friendly interfaces, while continuous monitoring and updates maintain accuracy as practices evolve.

This framework outlines a comprehensive pipeline for developing and implementing AI models in LT. The process spans seven stages: (1) defining clinical objectives and ensuring adherence to institutional review board guidelines; (2) collecting and harmonizing diverse data types (e.g., EHR, imaging, genetic data); (3) developing models tailored to tabular, imaging, and survival data, with a focus on interpretability; (4) validating models at single-center, multicenter, and clinical trial levels; (5) deploying models into clinical workflows via EHR integration; (6) monitoring performance, bias, and fairness; and (7) adapting to new treatments, evolving practices, and policy changes. This framework ensures rigorous, ethical, and effective deployment of AI in LT.

To evaluate the current state of AI applications in LT, we conducted a narrative review of studies published between January 2021 and October 2024. The reviewed studies encompass a broad spectrum of the LT process, including pre-transplant risk assessment, donor liver evaluation, transplant oncology, graft survival prediction, overall survival prediction, immunosuppression management, and post-transplant risk assessment. Geographical analysis highlights research efforts concentrated in North America (USA and Canada), Asia (notably China and South Korea), and Europe (with contributions from Italy, Spain, and Germany), alongside Brazil in South America [Figure 2A]. Figure 2B presents the average citation count across different problem areas in LT, emphasizing the relative impact and interest in each domain. Furthermore, the publication distribution demonstrates varying levels of research activity across all categories, as illustrated in Figure 2C. Notably, studies related to hepatocellular carcinoma (HCC) achieved



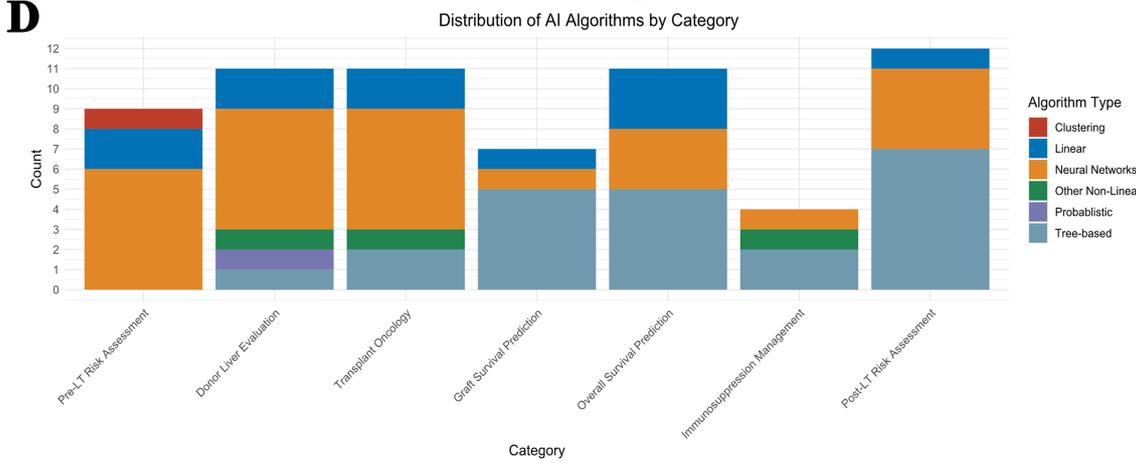
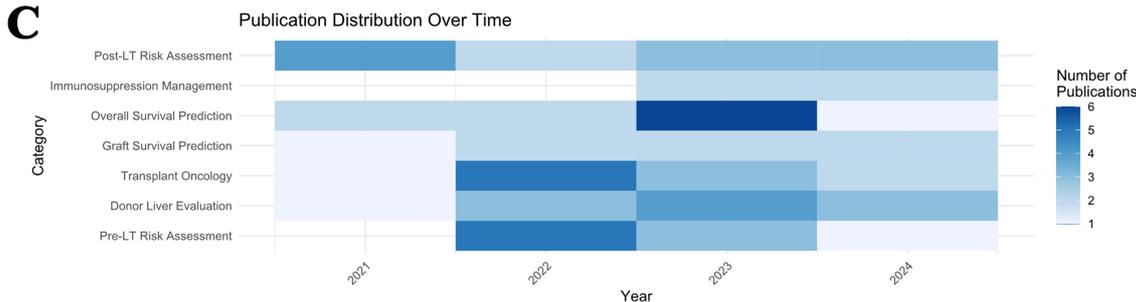
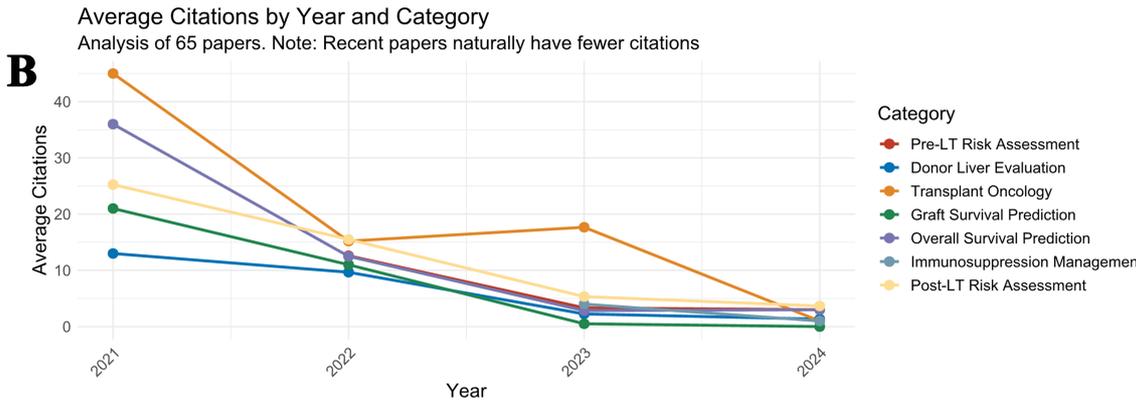
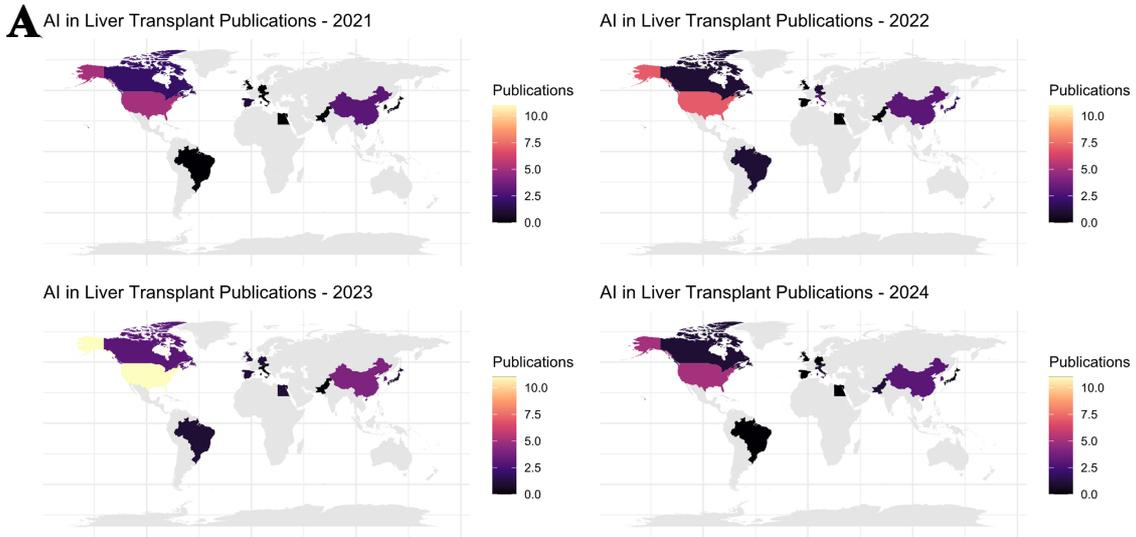
**Figure 1.** A systematic framework for AI development and deployment in LT. AI: Artificial intelligence; LT: liver transplantation.

the highest mean citation count, underscoring their prominence and influence within the field. The diversity of algorithmic approaches further underscores the multifaceted nature of AI applications: tree-based models are frequently employed for tabular clinical data, such as graft survival and overall survival, while neural networks and other non-linear models are increasingly applied to imaging and multimodal datasets, including donor liver assessments. [Figure 2D](#) further highlights the distribution of algorithms within each problem area, reflecting the tailored adaptation of AI methodologies to address specific challenges in LT.

This narrative review synthesizes advancements in AI-driven approaches across the LT continuum, evaluating their clinical relevance and feasibility. Critical considerations such as interpretability, fairness, and clinical integration are discussed, alongside future directions for advancing AI in LT. This comprehensive analysis underscores AI's transformative potential while highlighting the challenges that must be addressed for its ethical, equitable, and effective implementation.

## METHODS

Our study was designed as a narrative review with a structured search methodology to provide a comprehensive overview of AI applications in adult LT. While we employed some structured elements typically associated with systematic reviews (defined search strategy, inclusion criteria, and screening process), our primary aim was not to assess comparative effectiveness or generate pooled effect estimates but to map the landscape of AI innovations across the transplant continuum. We conducted a structured search in PubMed for original research articles published between January 2021 and October 2024, using keywords related to “artificial intelligence”, “machine learning”, “deep learning”, “liver transplantation”, and “hepatic transplantation”, combined with Boolean operators [[Supplementary Figure 1](#)]. Review articles, meta-analyses, case reports, editorials, and letters were excluded using publication-type filters. Only English-language studies involving human adults were retained. After running the search, we identified 100 potential articles. Titles and abstracts were screened for relevance (i.e., a clear focus on AI in LT), appropriate methodology (original, quantitative or mixed-methods research), and population criteria (adult patients). During screening, 35 articles were excluded for the following reasons: 12 articles were excluded



**Figure 2.** Trends and distribution of AI in liver transplant research (2021-2024). Multi-panel visualization depicting AI applications in LT research: (A) Geographic distribution of publications highlights the global contribution to AI in LT, led by the United States; (B) Average citations by year and category show increasing impact, particularly in Transplant Oncology (HCC recurrence prediction); (C) Temporal trends in research focus are shown in a heatmap depicting the distribution of publications across categories over time; and (D) algorithm distribution analysis reveals a predominance of neural networks and tree-based methods. Other non-linear approaches include k-Top Scoring Pairs and MARS. AI: Artificial intelligence; LT: liver transplantation; HCC: hepatocellular carcinoma; MARS: multivariate adaptive regression splines.

for lack of relevance (i.e., they did not pertain to AI in LT); 10 articles were excluded due to methodological issues, such as being review papers or having statistical flaws; 7 were excluded for focusing solely on pediatric or infant populations; 3 were excluded for limited or insufficient sample size; and 3 were excluded for reasons classified as other (research letters, inaccessible full text, or retracted articles). Ultimately, 65 articles met all inclusion criteria and were included in the review [Supplementary Figure 1]. Given the narrative scope of this review and the heterogeneity of the included studies (spanning diverse prediction tasks, algorithms, and clinical contexts), we did not perform a formal risk-of-bias assessment. Instead, our analysis focused on identifying trends, innovations, and challenges in applying AI throughout the LT process. Studies with substantial methodological limitations were excluded during the initial screening phase.

## OVERVIEW OF CURRENT AI APPROACHES

A broad array of AI methodologies has been employed in LT, as summarized in Table 1. The most frequently used approaches include tree-based models and neural networks. Tree-based models<sup>[4-8]</sup> mimic clinical decision making by hierarchically splitting data, efficiently handling categorical and continuous variables common in LT datasets. Neural networks excel at capturing complex non-linear relationships in diverse data types, which is particularly effective for medical imaging tasks such as organ segmentation<sup>[9-12]</sup>. Linear models, such as logistic and linear regression, provide interpretable results for more complex approaches.

## MODEL EVALUATION AND VALIDATION

Model assessment in LT relies on well-established performance metrics and validation strategies. Cross-validation and independent test datasets are used to validate model performance on unseen data. Cross-validation is particularly beneficial for smaller datasets, as it divides the data into subsets (“folds”) and iteratively uses each subset for training and testing. This process ensures that performance estimates are robust and not overly dependent on a single train-test split. For classification tasks, accuracy measures the proportion of correctly identified outcomes, while the area under the curve (AUC) is more appropriate for imbalanced datasets as it better reflects a model’s discriminative ability. For imaging segmentation tasks such as CT or MRI, the Dice coefficient quantifies the overlap between model predictions and expert-labeled boundaries. Finally, for continuous predictions, mean absolute error (MAE) indicates the average magnitude of errors, while root mean square error (RMSE) places greater emphasis on larger discrepancies.

## AI IN PRE-TRANSPLANT RISK ASSESSMENT

Traditional pre-transplant risk assessment tools provide value but may lack precision for individualized predictions. They rely on static variables and generalized scoring systems, which fail to capture nuanced, patient-specific interactions and complex patterns influencing outcomes. AI approaches are emerging as powerful complementary tools for risk stratification [Figure 3].

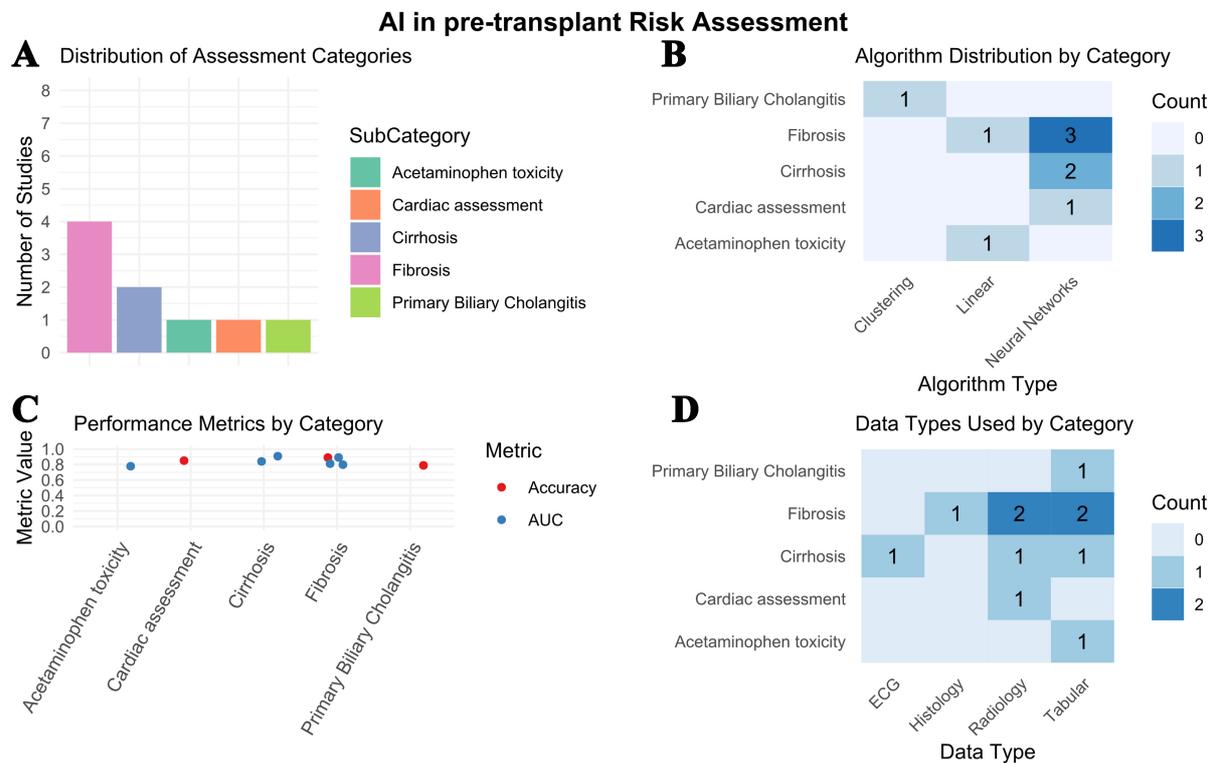
**Table 1. Overview of commonly used AI algorithms, their principles, main advantages, disadvantages, and potential applications in LT research and practice**

Algorithm	Principle	Advantages	Disadvantages	Applications
Logistic/linear/Cox regression	Linear: Fits a linear function to predict outcomes Cox: Semi-parametric approach modeling hazard rates over time (survival analysis)	- Clear interpretability (coefficients) - Fast to train and easy to implement	- Limited to linear relationships - Can underperform on high-dimensional data	- Predicting graft failure and survival post-LT
Tree-based (random forest, XGBoost)	Ensemble of decision trees (random forest uses multiple bagged trees; XGBoost uses boosting to iteratively improve weak learners)	- Can capture non-linearities and interactions - Often good performance out of the box	- Prone to overfitting if hyperparameters are not tuned - Less transparent than linear models (though easier to interpret than neural networks)	- Fibrosis staging from imaging - Survival or complication prediction using tabular registry data
Neural networks (CNN, MLP)	Deep architectures with multiple hidden layers; learn data representations through weighted connections updated by backpropagation	- Highly flexible; capture complex, non-linear relationships - Can handle diverse data types (images, text, labs)	- Requires large labeled datasets - Prone to overfitting - Less interpretable ("black box") - High computational cost	- Liver fibrosis staging from imaging - Liver segmentation for volume estimation - Steatosis assessment from histology
Transformers	Self-attention mechanism to capture contextual relationships in sequence data (time series, text)	- State-of-the-art in text data - Transfer learning via large-scale pretraining	- High computational cost - Interpretability hinges on attention maps - Often require large data to fine-tune effectively	- Steatosis grading from histology - Longitudinal complication forecasting
Unsupervised learning (clustering, autoencoders)	Learns patterns from unlabeled data by finding intrinsic structures (clusters) or compressed representations (autoencoders)	- Does not need labeled data - Discovers hidden phenotypes or groupings - Can be used for outlier detection	- Interpretability can be challenging - Results can be less intuitive without clinical input	- Phenotype discovery (new subtypes of liver disease) - Feature extraction for prediction tasks - Donor or recipient subgrouping

AI: Artificial intelligence; LT: liver transplantation; XGBoost: eXtreme gradient boosting; CNN: convolutional neural network; MLP: multilayer perceptron.

Liver fibrosis staging is a critical component of pre-transplant evaluation, as it informs predictions of liver disease progression and the need for transplant. As shown in [Figure 3A](#), liver fibrosis staging is the most extensively studied application of AI in pre-transplant assessment. Neural networks are prominently used models across multiple imaging modalities [[Figure 3B](#)]. For instance, Wada *et al.* developed a spectral CT technique for measuring extracellular volume fraction as a fibrosis marker with an AUC of 0.89<sup>[13]</sup>. Yu *et al.* focused on histological analysis, with a CNN-based UNet model performing portal tract segmentation from whole-slide images, achieving high correlations with clinical Scheuer staging<sup>[14]</sup>. Ahn *et al.* trained a convolutional neural network (CNN) on electrocardiogram images to predict cirrhosis<sup>[15]</sup> with an AUC of 0.908, while Mazumder *et al.* developed a CT-based liver segmentation tool to predict cirrhosis with an AUC of 0.84<sup>[16]</sup>. AI fibrosis assessment helps surgical teams anticipate portal hypertension severity, which informs decisions about vascular management strategies and the potential need for additional considerations during the hepatectomy phase.

While this section focuses on pre-transplant applications, it is worth noting that fibrosis staging also plays a critical role in post-transplant settings. Graft fibrosis after LT also has significant implications for graft and patient survival. Azhie *et al.* developed a long short-term memory (LSTM)-based deep learning model combining longitudinal clinical and laboratory data to non-invasively predict graft fibrosis<sup>[17]</sup> with an AUC of 0.798. Similarly, Qazi Arisar *et al.* utilized radiomic and clinical features using a linear model (lasso) to predict graft fibrosis<sup>[18]</sup> with an AUC of 0.811.



**Figure 3.** Overview of machine learning applications in pre-liver transplant risk assessment (2021-2024). (A) Distribution of studies across subcategories (acetaminophen toxicity, cardiac assessment, cirrhosis, fibrosis, primary biliary cholangitis); (B) Algorithm types used by category (clustering, linear methods, neural networks), with neural networks showing predominance particularly in image-based studies; (C) Performance metrics comparison across categories using accuracy and AUC; and (D) Distribution of data types utilized (ECG, histology, radiology, tabular), highlighting the increasing adoption of imaging-based approaches. AUC: Area under the curve; ECG: electrocardiogram.

AI applications targeting specific liver diseases have also gained traction. Gerussi *et al.* applied clustering algorithms using clinical variables to predict death or the need for transplant in primary biliary cholangitis, achieving an accuracy of 0.789<sup>[19]</sup>. Umbaugh *et al.* developed a logistic regression model to predict outcomes in acetaminophen-induced acute liver failure using clinical variables, with an AUC of 0.778<sup>[20]</sup>.

Cardiac assessment is another critical aspect of pre-transplant evaluation. Schuessler *et al.* used a neural network to analyze coronary CT angiography-derived fractional flow reserve in transplant candidates for the assessment of coronary artery disease (CAD), achieving an accuracy of 0.85, demonstrating promising potential for non-invasive cardiac evaluation<sup>[21]</sup>. AI-detected coronary disease informs anesthetic monitoring choices and guides surgical teams to modify hemodynamic management during critical phases of transplantation.

Performance metrics across these AI applications demonstrate promising outcomes [Figure 3C]. [Supplementary Table 1](#) offers a detailed overview of studies on AI in pre-transplant risk assessment, summarizing prediction focus, variable modalities, algorithms, study designs, interpretation methods, limitations, sample sizes, and key performance metrics.

## AI IN DONOR LIVER ASSESSMENT

Due to the shortage of donor organs in LT and the need to increase the utilization of extended-criteria donors<sup>[22]</sup>, there is a growing need for AI-based tools to improve organ assessment accuracy. Current AI applications primarily focus on steatosis assessment, followed by liver volume estimation and biliary tract visualization [Figure 4A].

Steatosis assessment is a critical determinant of graft survival, as organs with significant macrovesicular steatosis (> 30%) are associated with higher risks of primary non-function and early allograft dysfunction<sup>[23]</sup>. Traditional pathologist assessments during deceased donor evaluations face considerable challenges, including inter-observer variability and time constraints. Real-time AI steatosis assessment directly impacts surgical decision making during procurement, helping teams determine organ suitability and anticipate post-transplant management needs.

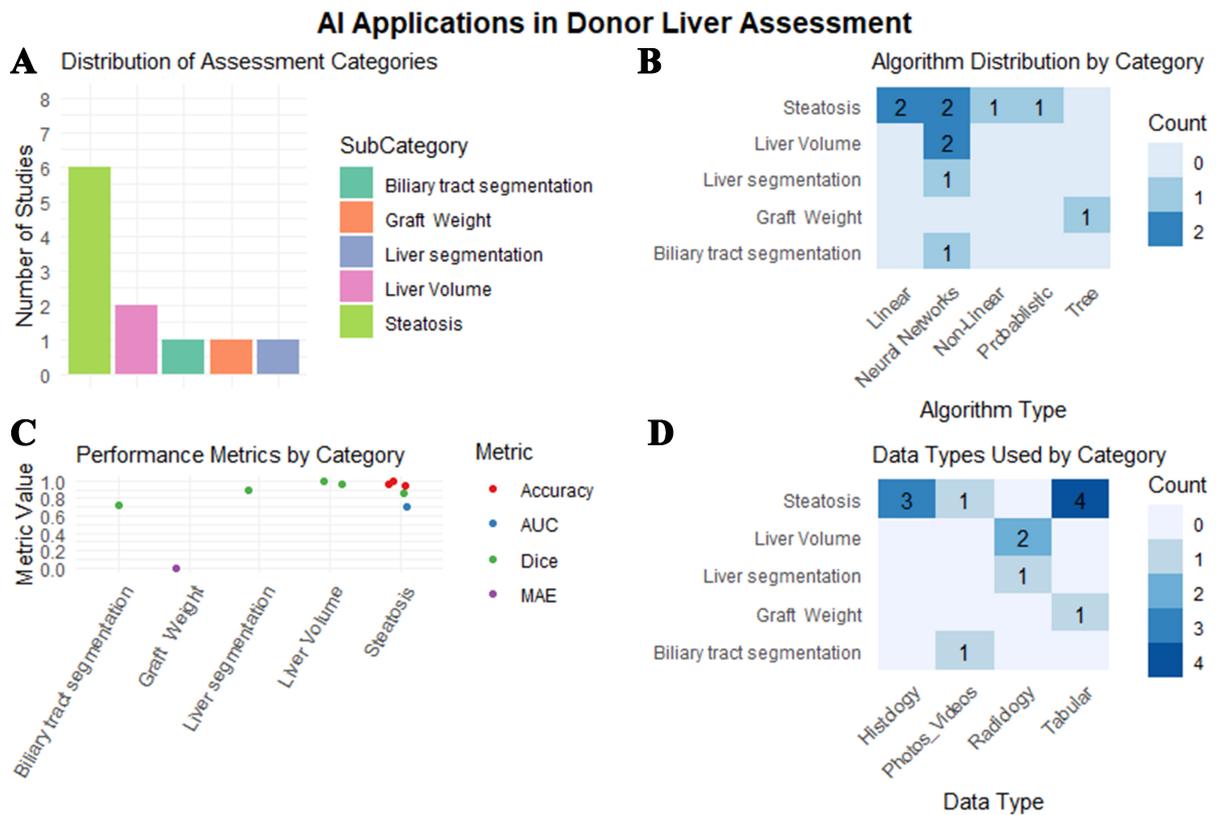
Pérez-Sanz *et al.* utilized a Naïve Bayes classifier to process Sudan-stained slides, achieving an accuracy of 0.99<sup>[24]</sup>. Similarly, Tang *et al.* employed transformer-based models for grading steatosis during frozen section analysis, achieving an accuracy of 0.96<sup>[25]</sup>. However, transformer-based models are slower on real-time inference compared to the CNN-based approaches. Gambella *et al.* developed a CNN model for histological evaluation of donor liver steatosis, achieving a Dice score of 0.85<sup>[26]</sup>.

Frey *et al.* used Logistic Lasso models to predict macrosteatosis (> 30%) in deceased donor livers using data from the united network for organ sharing (UNOS) database, achieving an AUC of 0.71<sup>[27]</sup>. Cherchi *et al.* applied a kernel-based K<sup>\*</sup> algorithm to predict < 30% hepatic steatosis, achieving an accuracy of 0.94, demonstrating the utility of AI for donor liver quality assessment<sup>[28]</sup>. Lim *et al.* developed a logistic regression model to predict hepatic steatosis (> 5%) in donor livers using clinical variables<sup>[29]</sup>.

Volume estimation accuracy is crucial for surgical safety and recipient outcomes, particularly in living donor LT. Jeong *et al.* developed deep attention LSTM U-Net (DALU-Net), which achieved a 99.6% Dice score in determining safe resection volumes<sup>[30]</sup>. Yang *et al.* developed transformer-based UNETR models, achieving a Dice score of 0.959 in right lobe graft weight estimation<sup>[31]</sup>. Giglio *et al.* applied a quantile random forest model to predict graft weight, reporting a MAE of 0.103 across multicenter datasets<sup>[32]</sup>. Similarly, Kazami *et al.* introduced a two-step AI algorithm for liver segmentation that automates anatomic virtual hepatectomy, achieving high Dice coefficients (0.71-0.95) for hemilivers, sectors, and Couinaud's segments<sup>[33]</sup>.

AI has also made significant advances in intraoperative guidance through biliary tract visualization. Oh *et al.* developed real-time segmentation models using DeepLabV3+ in bile duct identification during laparoscopic donor hepatectomy with a Dice score of 0.728<sup>[34]</sup>. These segmentation and volume estimation tools may help prevent complications such as small-for-size syndrome, optimizing graft selection and improving patient outcomes.

Performance metrics across these AI applications underscore their transformative potential in donor liver assessment. As shown in Figure 4, high AUCs, Dice scores, and low error rates reported by studies like those of Pérez-Sanz *et al.*<sup>[24]</sup>, Lim *et al.*<sup>[29]</sup>, Tang *et al.*<sup>[25]</sup>, and Jeong *et al.*<sup>[30]</sup> demonstrate the capacity of AI to improve accuracy, reduce variability, and streamline decision making in both preoperative and intraoperative settings. Supplementary Table 2, Figure 4B and D further illustrate the diversity of algorithms and data types employed in these studies.



**Figure 4.** Overview of machine learning applications in donor liver assessment (2021-2024). (A) Distribution of studies across subcategories (biliary tract segmentation, graft weight, liver segmentation, liver volume, steatosis), with steatosis being the most studied area; (B) Algorithm types used by category (linear, neural networks, non-linear, probabilistic, tree-based methods), with neural networks dominating due to increased use of images; (C) Performance metrics comparison across categories using multiple measures (Accuracy, AUC, Dice coefficient, MAE); and (D) Distribution of data types utilized (histology, liver photos/videos, radiology, tabular), demonstrating heavy reliance on histological and radiological data, particularly for steatosis assessment. AUC: Area under the curve; MAE: mean absolute error.

## AI IN TRANSPLANT ONCOLOGY

HCC’s complex tumor biology and heterogeneous presentation make standardized decision making difficult<sup>[35]</sup>. AI looks promising in addressing HCC management by addressing critical challenges such as heterogeneous tumor biology, limitations of size-based eligibility criteria like the Milan criteria, and the complexity of processing vast imaging, clinical, and molecular datasets. AI enhances decision making in both pre- and post-transplant phases.

In pre-transplant assessment, Huang *et al.* introduced the denoised local and non-local features fusion network (DLNLF-Net), which used MRI images to grade HCC malignancy with an AUC of 0.95, highlighting the potential of deep learning in refining patient selection<sup>[36]</sup>. Similarly, Kwong *et al.* employed Cox regression on the UNOS database to predict waitlist dropout at 3, 6, and 12 months post-listing, achieving a C-index of 0.74<sup>[37]</sup>.

For post-transplant recurrence prediction, He *et al.* developed a multimodal deep learning model combining CNNs and multilayer perceptrons (MLPs) to predict recurrence with an AUC of 0.87, aiming to expand LT options beyond traditional size-based criteria to include patients with larger tumor burdens (> 5 cm)<sup>[38]</sup>. Liu *et al.* focused on CNN-based classification with weak supervision using histological slides

and clinical data, enhancing prognostic stratification<sup>[39]</sup>.

Ivanics *et al.* created the Toronto Post-Liver Transplantation HCC Recurrence Calculator using regularized Cox regression with a C-index of 0.75<sup>[40]</sup>. Liu *et al.* integrated transcriptome and exome analyses using the k-Top Scoring Pairs algorithm to predict recurrence, achieving a C-index of 0.81, paving the way for personalized risk assessments<sup>[41]</sup>. Tran *et al.* developed the RELAPSE score using random survival forest analysis in a large, multicenter cohort of 6,141 patients<sup>[42]</sup> with a C-index of 0.81 for post-LT HCC recurrence. The external validation of this model in an external European cohort represented a significant step toward overcoming the limitations of earlier single-center studies, though challenges in capturing granular pre-transplant treatment data persisted. Qu *et al.* utilized CNNs with attention mechanisms to process histology data predicting recurrence-free survival, offering visual explanations through attention heatmaps with an AUC of 0.85<sup>[43]</sup>. To *et al.* analyzed gene expression profiles using CNNs for HCC recurrence post-LT<sup>[44]</sup>. Iseke *et al.* developed an integrated approach using CNN and eXtreme gradient boosting (XGBoost) to predict HCC recurrence 1-6 years post-transplant<sup>[45]</sup>. Altaf *et al.* created a MLP neural network focusing on 5-year recurrence-free survival prediction with an AUC of 0.77<sup>[46]</sup>.

[Supplementary Table 3](#) complements [Figure 5](#) by summarizing individual studies, detailing prediction tasks, algorithm types, performance metrics, and dataset characteristics, providing a deeper understanding of AI's contributions and limitations in transplant oncology.

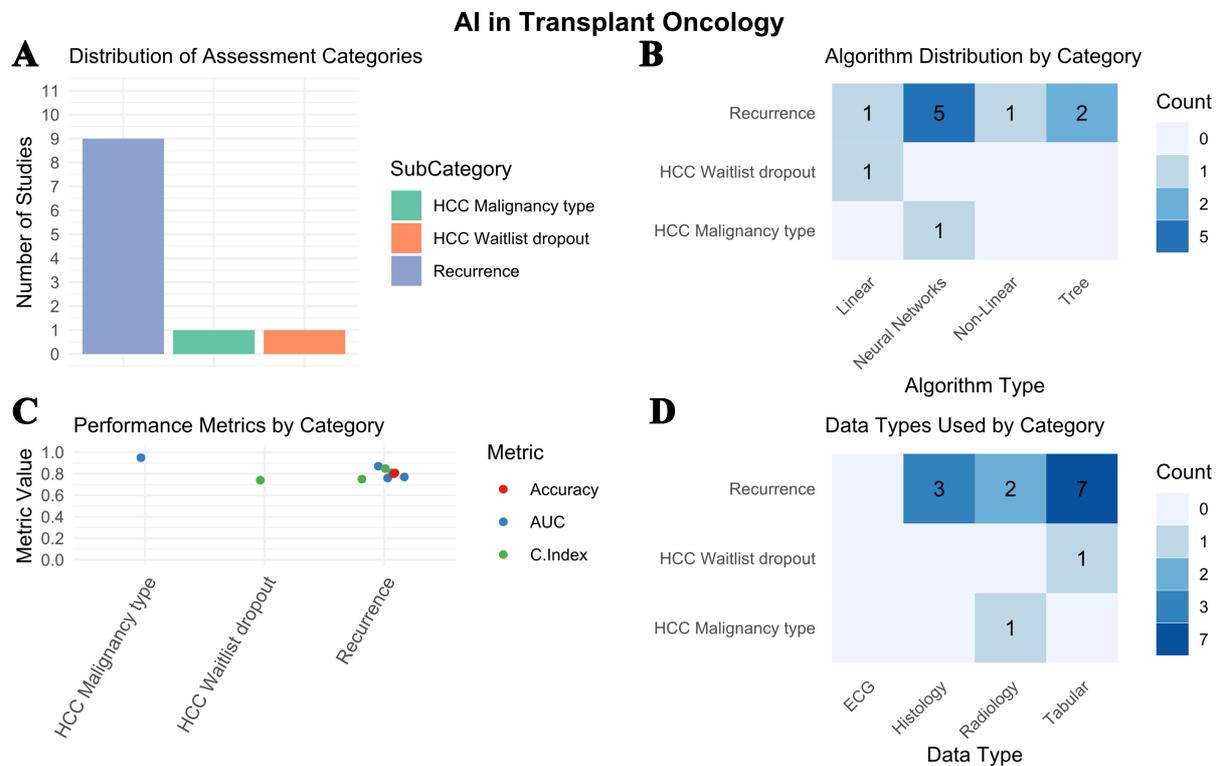
## AI FOR GRAFT SURVIVAL PREDICTION

Graft survival prediction is of critical importance in LT, given its potential to influence organ allocation [[Figure 6](#)]. Research has primarily focused on general graft survival prediction, with some studies addressing specific complications like graft-versus-host disease (GVHD, [Figure 6A](#)). Guijo-Rubio *et al.* analyzed the UNOS database with 39,189 donor-recipient pairs, demonstrating that traditional logistic regression models could outperform more complex machine learning approaches for 5-year graft survival prediction (AUC = 0.654)<sup>[47]</sup>. This finding highlights the value of simplicity in certain contexts.

For shorter-term predictions, Yanagawa *et al.* developed a LightGBM-based model for 90-day graft failure<sup>[8]</sup> with an AUC of 0.7. Recent studies have also explored post-transplant complications, such as Cooper *et al.*'s decision tree-based model for predicting the rare complication of GVHD using data from 2,013 cases<sup>[48]</sup> with an AUC of 0.86. Addressing fairness in organ allocation, Ding *et al.* analyzed the UNOS standard transplant analysis and research (STAR) file using a novel two-step debiasing strategy that combined tree-based models and neural networks to ensure equitable predictions across different demographic groups<sup>[49]</sup> with an AUC of 0.79. Lin *et al.* integrated proteomics and metabolomics data to predict early allograft dysfunction<sup>[50]</sup> (AUC = 0.833), while Bambha *et al.* developed a random forest survival model specifically to optimize outcomes in non-directed living liver donor transplants<sup>[51]</sup> with a C-index of 0.63. Zalba Etayo *et al.* focused on 1-year graft survival incorporating comorbidities, using an artificial neural network (C-index = 0.745)<sup>[52]</sup>. Interestingly, all graft survival studies used tabular data of clinical features [[Figure 6D](#)]. Future studies can focus on integrating multimodal data types. [Supplementary Table 4](#) provides a detailed summary of these studies, including sample size for each study, performance metrics, and study design.

## AI FOR OVERALL SURVIVAL PREDICTION

AI has the potential to significantly advance post-LT survival prediction, encompassing both short-term and long-term outcomes through diverse methodologies. Overall survival prediction constitutes the largest category of studies, with additional research targeting specific populations, such as patients with sarcopenia or diabetes mellitus [[Figure 7A](#)].

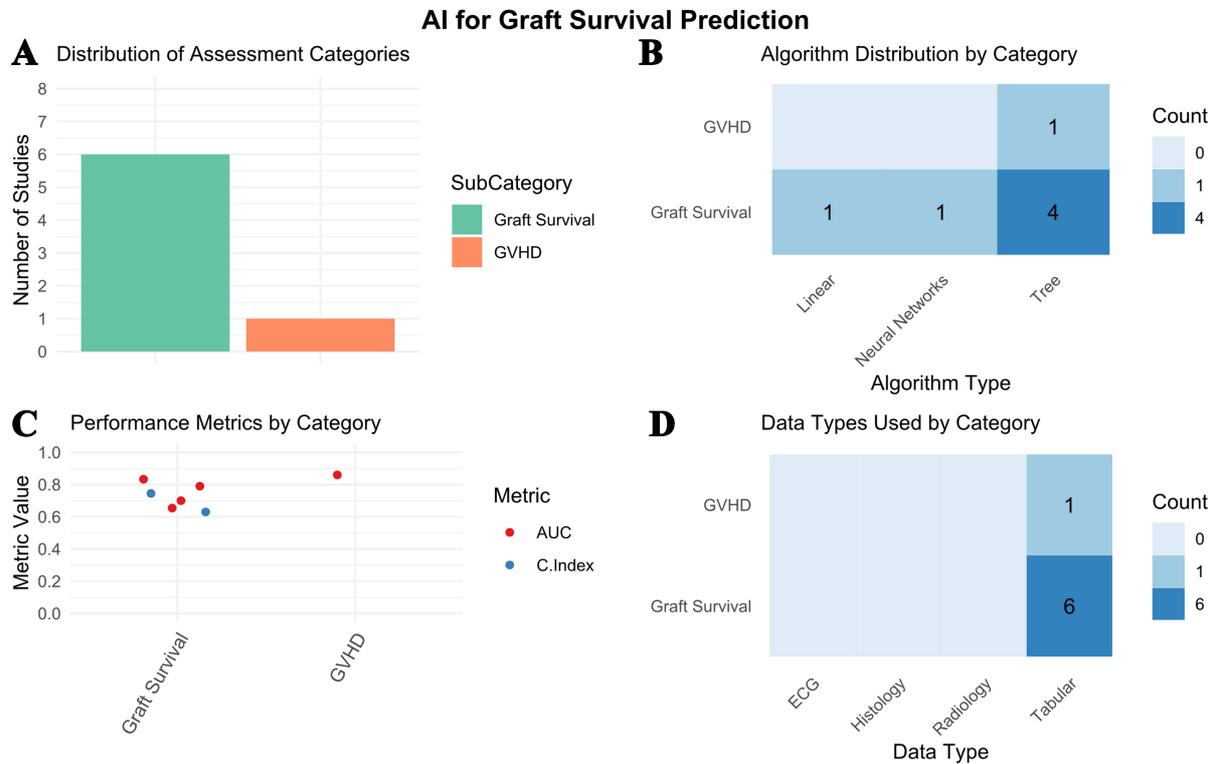


**Figure 5.** Overview of machine learning applications in transplant oncology (2021-2024). (A) Distribution of studies across subcategories (HCC Malignancy type, HCC Waitlist dropout, Recurrence), with recurrence being the predominant focus area; (B) Algorithm types used by category (linear, neural networks, non-linear, tree-based methods), showing strong preference for neural networks particularly in recurrence prediction; (C) High performance metrics comparison across categories using multiple measures (Accuracy, AUC, C-index); and (D) Distribution of data types utilized (histology, radiology, tabular), highlighting extensive use of tabular data especially in recurrence studies along with increase in use of radiology and histology data. HCC: Hepatocellular carcinoma; AUC: area under the curve.

For short-term survival, AI models have shown great promise. Yu *et al.* utilized random forest algorithms to predict one-year post-LT survival, achieving an AUC of 0.81, outperforming traditional MELD-based methods<sup>[4]</sup>. Similarly, Yang *et al.* applied random forest models to predict 90-day survival in patients with acute-on-chronic liver failure (ACLF) with an AUC of 0.94<sup>[53]</sup>. Combining machine learning with clinical expertise, Ge *et al.* used gradient boosting machines enhanced by physician input for one-year survival predictions in ACLF patients with an AUC of 0.719<sup>[5]</sup>. [Figure 7](#) shows the high performance of both models.

For longer-term survival, AI approaches have also demonstrated potential. Yasodhara *et al.* employed a Cox proportional hazards model to predict mortality in diabetic LT recipients, identifying hypertension and renal dysfunction as key predictors of survival with a C-index of 0.61<sup>[54]</sup>. Transformer-based deep learning algorithms have expanded predictive capabilities. For example, Nitski *et al.* developed a transformer model to predict cause-specific mortality due to complications such as cancer, cardiovascular events, infection, and graft failure<sup>[11]</sup>.

On a global scale, Ivanics *et al.* conducted a multi-national study using ridge regression, reporting AUCs ranging from 0.64 to 0.74 across various registries, while addressing limitations in cross-country model transferability<sup>[55]</sup>. Park *et al.* leveraged 3D CNN models to analyze CT scans for sarcopenia assessment and survival prediction, demonstrating the diverse data types being utilized [[Figure 7D](#)]<sup>[56]</sup>. Liu *et al.* combined



**Figure 6.** Overview of machine learning applications in graft survival prediction (2021-2024). (A) Distribution of studies across subcategories (graft survival, GVHD), with graft survival being the primary focus area; (B) Algorithm types used by category (linear, neural networks, tree-based methods), showing strong preference for tree-based methods due to the use of only tabular data; (C) Performance metrics comparison across categories using AUC and C-index, with both categories showing promising results; and (D) Distribution of data types utilized showing only use of tabular data for both graft survival and GVHD predictions. GVHD: Graft-versus-host disease; AUC: area under the curve.

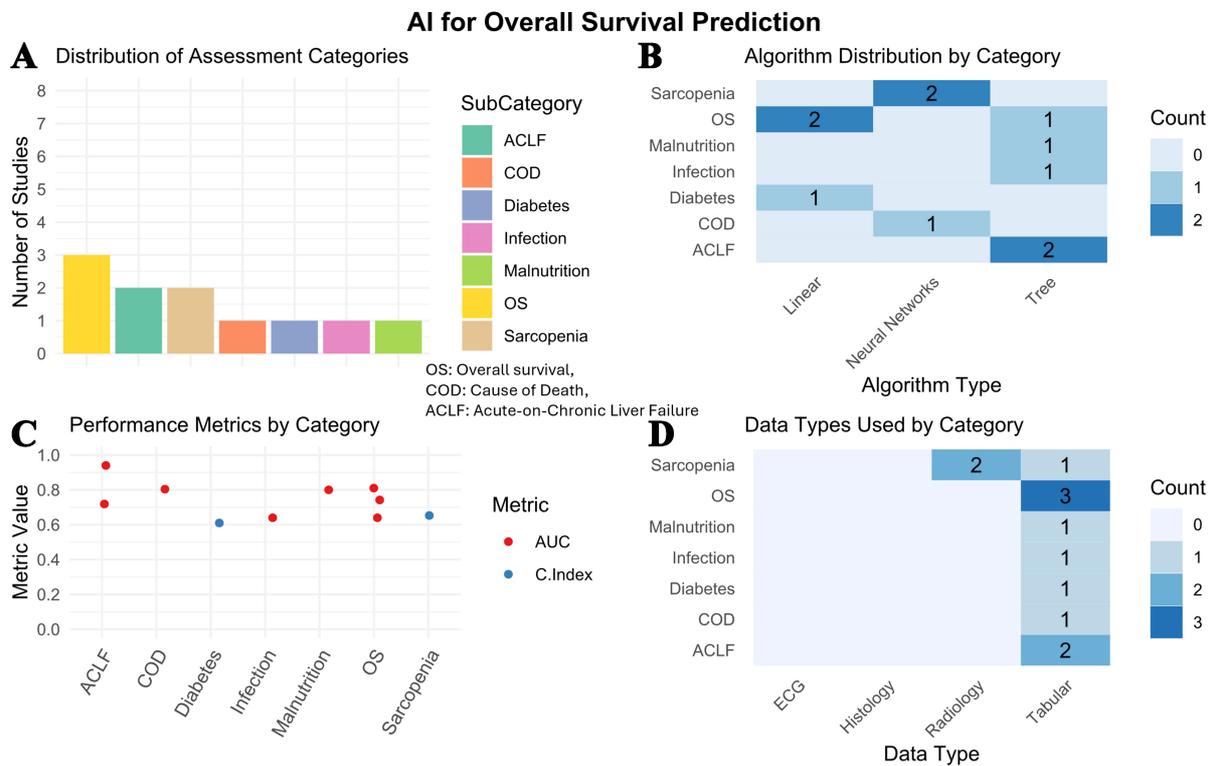
CT imaging with tabular data using DeepSurv and autoencoder models to predict survival post-LT for patients with HCC and sarcopenia, achieving a C-index of 0.653<sup>[57]</sup>.

In addition to these studies, Fonseca *et al.* used random forest models to predict one-year post-transplant mortality by factoring in malnutrition during the waitlist period, achieving an AUC of 0.8<sup>[58]</sup>. Ding *et al.* applied multi-task gradient boosting models to analyze cause-of-death data post-transplant, achieving an AUC of 0.64 while addressing fairness<sup>[59]</sup>. Rogers *et al.* developed an interpretable survival tree algorithm (SurvCART) for patient-specific mortality estimation, reporting an AUC of 0.74<sup>[60]</sup>.

Figure 7 and Supplementary Table 5 provide a detailed summary of these studies.

## AI IN IMMUNOSUPPRESSION MANAGEMENT

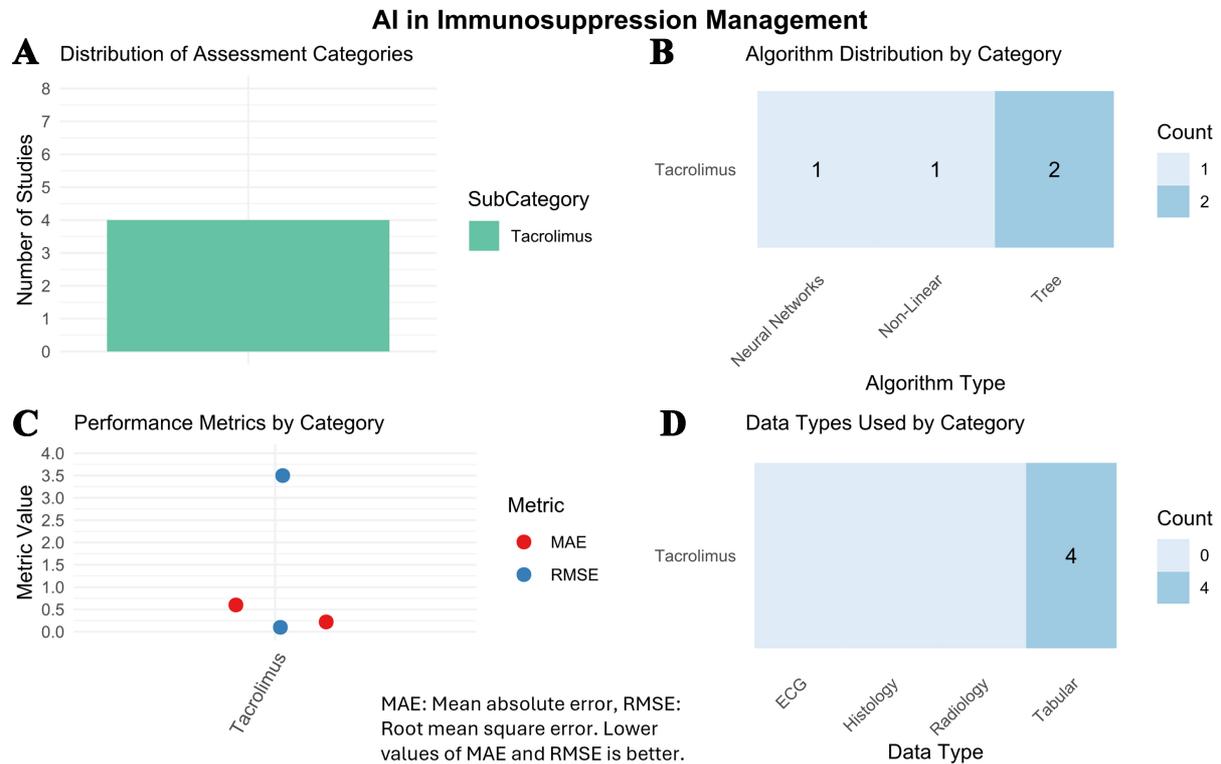
Precise tacrolimus management is vital in LT due to its narrow therapeutic window. Subtherapeutic levels can result in graft rejection, while overexposure increases the risk of nephrotoxicity, neurotoxicity, and infections. This variability in drug metabolism among individuals underscores the need for accurate dosing strategies and monitoring. Recent advancements in AI have shown significant promise in addressing these challenges, improving both prediction accuracy and clinical utility [Figure 8].



**Figure 7.** Overview of machine learning applications in overall survival prediction (2021-2024). This figure illustrates the diverse applications of machine learning in predicting mortality and survival outcomes related to LT and associated conditions. (A) Distribution of studies across prediction categories, where each category represents specific survival problems such as mortality prediction in diabetic patients, survival in ACLF patients, cause of death analysis, overall survival at various time points, and post-transplant outcomes in patients with complications like malnutrition, infection, and sarcopenia; (B) Utilization of different machine learning algorithms (linear models, neural networks, and tree-based methods) across prediction tasks; (C) Model performance comparison using AUC and C-index metrics; (D) Data types employed across studies, showing the predominance of tabular data and specialized use of radiological imaging for specific conditions, such as in sarcopenia assessment. LT: Liver transplantation; ACLF: acute-on-chronic liver failure; AUC: area under the curve.

AI approaches have ranged from traditional methods to neural networks [Figure 8A]. For instance, Ponthier *et al.* utilized multivariate adaptive regression splines (MARS) to estimate the tacrolimus AUC using limited blood concentration data points in a multicenter study involving 161 patients, achieving a RMSE of 0.0986<sup>[61]</sup>. Similarly, Li *et al.* developed a hybrid model combining population pharmacokinetics with XGBoost in a cohort of 177 Chinese LT recipients<sup>[62]</sup>. Their approach integrated genetic polymorphism data, emphasizing key factors such as voriconazole co-administration and genetic variations in CYP3A41G<sup>+</sup> and CYP3A53<sup>+</sup>. Their model achieved a RMSE of 3.5.

Advancing these efforts, Du *et al.* employed MLP neural networks to analyze tacrolimus pharmacokinetics<sup>[63]</sup>. Their model, applied in a single-center study with 31 patients, reported an MAE of 0.6. Recently, Yoon *et al.* implemented LSTM networks in a multicenter study of 443 patients, achieving an MAE of 0.22<sup>[64]</sup>. Their findings showed practical clinical relevance, with patients receiving doses outside model suggestions experiencing longer ICU stays by an average of 2.5 days. LSTM models, particularly effective with time-series data, are well-suited for predicting tacrolimus concentrations and optimizing dose adjustments.



**Figure 8.** Applications of AI in immunosuppression management for LT (2021-2024). (A) Distribution of assessment categories shows 4 studies focused on tacrolimus dosing and monitoring; (B) Tree-based methods were slightly more prevalent with 2 studies; (C) Performance metrics (MAE and RMSE) demonstrating high reliability across applications (lower the better); and (D) Only tabular clinical data used for model development. LT: Liver transplantation; MAE: mean absolute error; RMSE: root mean square error.

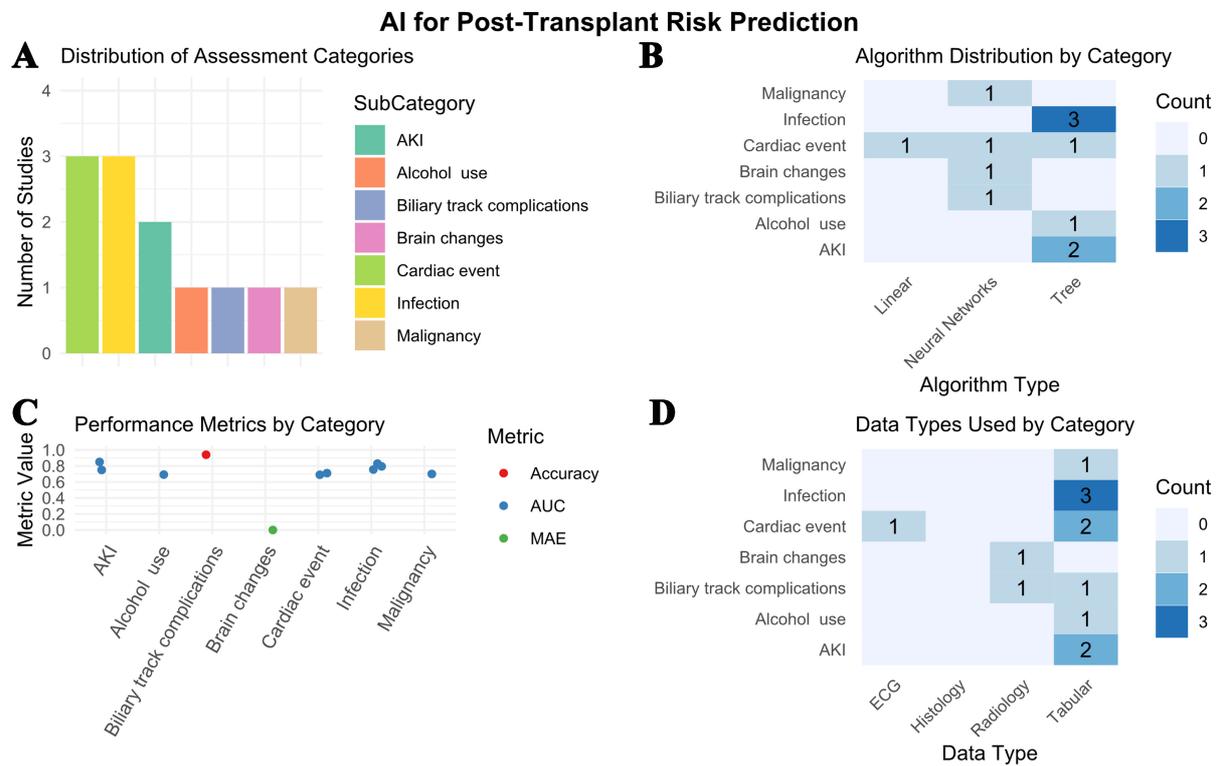
Supplementary Table 6 provides a detailed summary of these studies. Figure 8 illustrates the distribution of assessment categories, algorithm types, data types, and performance metrics used in these studies. All studies utilized tabular data for their predictions [Figure 8D], highlighting the prevalence of structured datasets in tacrolimus management research. Notably, Figure 8A underscores the limited number of studies in this area.

## AI FOR POST-TRANSPLANT COMPLICATIONS PREDICTION

The early prediction of complications following LT is critical for improving patient outcomes. Timely identification enables the implementation of preventive interventions and targeted management strategies.

Acute kidney injury (AKI) remains one of the most common early complications post-LT, as shown in Figure 9A. Zhang *et al.* developed a gradient boosting machine learning model, achieving an AUC of 0.75 for AKI prediction<sup>[65]</sup>, while He *et al.* applied a random forest model with an AUC of 0.85<sup>[6]</sup>.

AI applications targeting post-LT infections have demonstrated significant advancements. Figure 9B highlights that infections are another well-studied area where tree-based models, such as random forests and XGBoost, dominate. For example, Chen *et al.* utilized an XGBoost model to predict pneumonia post-LT, achieving an AUC of 0.794<sup>[66]</sup>. Freire *et al.* addressed antimicrobial resistance in a multicenter retrospective study with a random forest model predicting carbapenem-resistant Enterobacterales carriage, achieving an AUC of 0.83<sup>[67]</sup>. Similarly, Chen *et al.* developed a random forest model to predict



**Figure 9.** Analysis of AI applications in post-liver transplant risk predictions (2021-2024). (A) Distribution of assessment categories shows cardiac events and infections as the most studied complications; (B) Algorithm distribution reveals a preference for neural networks and tree-based methods across complications; (C) Performance metrics demonstrate high accuracy/AUC across categories; (D) Data types utilized vary by complication, with tabular data being most common. Complications studied include AKI, infections, cardiac events, brain changes, alcohol use, biliary complications, and malignancy. AI: Artificial intelligence; AUC: area under the curve; AKI: acute kidney injury.

postoperative sepsis within 7 days after LT, achieving an AUC of 0.75<sup>[68]</sup>.

Cardiac events are another critical post-transplant complication, as shown in [Figure 9A](#). Jain *et al.* used XGBoost to predict major adverse cardiovascular events (MACE), achieving an AUC of 0.71<sup>[7]</sup>. Jang *et al.* extended this approach, employing XGBoost to predict MACE events post-LT using clinical variables and echocardiogram readings for mitral annular calcification<sup>[69]</sup>. Neural networks, particularly CNNs, are increasingly used for cardiac risk prediction, as highlighted in [Figure 9B](#), with studies like Zaver *et al.*'s analyzing electrocardiogram (ECG) data to predict atrial fibrillation and low ejection fraction post-LT<sup>[70]</sup>, achieving an AUC of 0.69. This AI risk score enables surgical teams to prepare for potential cardiac complications by adjusting medication protocols and ensuring appropriate monitoring.

For biliary complications, which affect up to 30% of LT recipients, Fodor *et al.* developed CNNs leveraging hyperspectral imaging for predictive analysis<sup>[71]</sup>. Cheng *et al.* employed CNNs to monitor longitudinal changes in brain structural patterns before and 1, 3, and 6 months after transplant<sup>[72]</sup>. Additionally, Lee *et al.* created an XGBoost model using psychosocial profiles from a multicenter cohort to predict harmful alcohol use post-LT<sup>[73]</sup>. Similarly, transformer-based deep learning models are being used to predict broader complications like malignancy, diabetes, rejection, and infection, demonstrating high performance while addressing demographic fairness<sup>[10]</sup>. Their model addressed demographic fairness, reducing task discrepancies by 39% while maintaining high predictive accuracy.

Figure 9C shows the performance metrics for AI models across various categories. As shown in Figure 9D, tabular data are the predominant input for AI models, although imaging modalities like radiology and histology are increasingly being incorporated to improve prediction accuracy. AI-based prediction models for post-transplant complications have demonstrated significant potential, leveraging diverse data modalities and achieving high predictive performance across categories. Figure 9 underscores the breadth of AI applications and the diversity of approaches being explored in post-transplant risk prediction. Supplementary Table 7 provides a detailed summary of these studies.

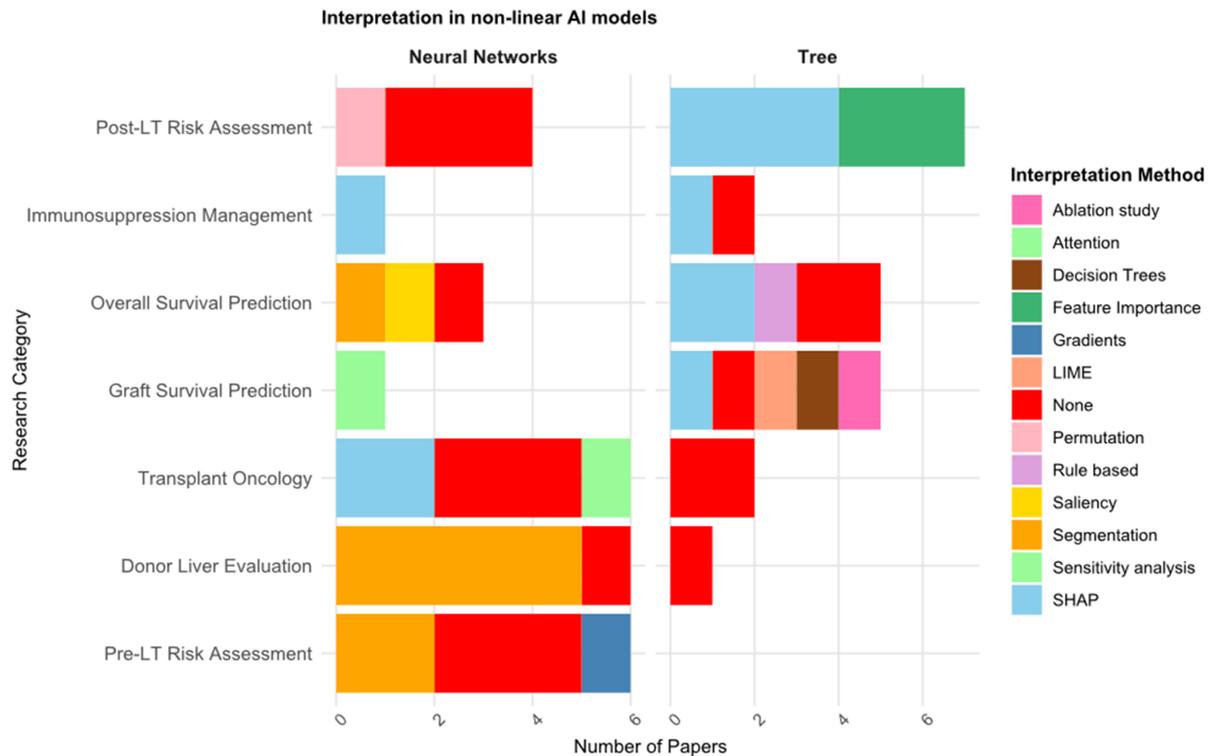
## INTERPRETABILITY IN AI FOR LT

Interpretability is the extent to which humans can understand and explain the decisions made by an AI model. In LT, interpretability is especially critical, as understanding AI predictions directly influences clinical decision making. Linear models, such as logistic and Cox regression, provide inherent transparency by assigning coefficients that quantify the influence of input variables on predictions. These feature coefficients represent the change in outcome associated with a one-unit change in a variable, assuming all other variables remain constant. While these models are computationally efficient and straightforward to interpret, they are limited by their assumptions of linearity, inability to capture feature interactions, and reliance on global importance measures that may not reflect individual case complexity.

Modern AI, as reviewed, often relies on complex models like neural networks and tree-based methods, which significantly improve predictive performance but often lack interpretability. Figure 10 highlights the interpretation approaches currently used in AI models for LT, particularly for neural networks and tree-based methods. Techniques such as SHapley Additive exPlanations (SHAP), rooted in game theory, quantify the contribution of each feature to an individual prediction but are computationally expensive<sup>[74]</sup>. Similarly, local interpretable model-agnostic explanations (LIME) creates simpler surrogate models to explain individual predictions<sup>[75]</sup>. Other methods, such as feature importance metrics (e.g., Gini index), identify key variables influencing outcomes. Advanced tools like attention mechanisms in transformer models and saliency maps for neural networks further enhance transparency by identifying parts of input data most critical for predictions. These tools collectively help bridge the gap between AI's predictive power and the need for explainability in healthcare. In addition, embedding known clinical constraints (e.g., MELD-based thresholds) or domain knowledge (e.g., Milan Criteria for HCC) within model outputs is another way to ensure the model's logic remains consistent with real-world transplant practices. Model outputs during both training and validation should also be routinely verified by physicians to align predictions with clinical expertise.

This review reveals notable gaps in the implementation of interpretability methods in LT AI research [Figure 10]. Among studies using neural networks, 11 out of 27 (40.7%) did not incorporate any interpretability approaches. Similarly, 7 out of 21 studies (33.3%) using tree-based models lacked interpretability methods. This represents a significant gap, as approximately one-third of studies employing complex AI models fail to address this crucial aspect.

This lack of interpretability is particularly concerning in the context of LT, where understanding model decisions is essential for clinical adoption and patient safety. The inconsistent adoption of interpretability methods across studies highlights the need for standardized reporting and implementation practices in LT research. Addressing these gaps will be crucial for ensuring that AI tools are transparent, reliable, and widely accepted in clinical practice. Encouraging the use of standardized frameworks (e.g., TRIPOD for transparent reporting) will also facilitate safe and equitable deployment of AI in LT<sup>[76]</sup>.



**Figure 10.** Distribution of interpretation methods across AI models in LT for non-linear neural network and tree-based approaches. Non-linear models, such as neural networks and tree-based approaches like random forests, present significant challenges in interpretability. This figure highlights the various methods employed to interpret these models. Neural network-based segmentation often serves as a self-explanatory approach, while a substantial proportion of studies (red) do not utilize any interpretation methods. In contrast, linear approaches, such as Linear regression or Cox regression, offer straightforward interpretability through coefficients. This figure underscores the gap in interpretability strategies for complex AI models in LT. AI: Artificial intelligence; LT: liver transplantation.

### ACHIEVING FAIRNESS IN AI FOR LT

Fairness in AI is defined by the TRIPOD guidelines as the “property of prediction models that do not discriminate against individuals or groups of individuals based on attributes such as age, race/ethnicity, sex/gender, or socioeconomic status”<sup>[76]</sup>. In the context of LT, fairness is particularly critical due to the scarcity of donor organs and the life-or-death consequences associated with allocation decisions. Traditional allocation systems, such as the MELD score, have demonstrated gender disparities<sup>[77]</sup>. As AI systems are increasingly proposed for transplant decision support, ensuring these systems do not perpetuate or amplify existing biases becomes paramount. Strauss *et al.* emphasize in their qualitative study that fairness in AI for LT requires careful attention to key factors, including transparency in AI development, a clear understanding of how algorithms process data, and the potential for AI to either mitigate or exacerbate biases<sup>[78]</sup>.

Despite its importance, fairness is addressed in only a small fraction (6%) of reviewed AI LT studies. However, two recent studies have made notable advancements in this area. Ding *et al.* proposed a comprehensive fairness framework incorporating a two-step debiasing strategy<sup>[49]</sup>. Their approach used knowledge distillation to handle dense features (e.g., lab values) and sparse categorical features (e.g., demographic data) and applied fairness constraints during model training. This strategy significantly reduced demographic biases, reducing gaps in positive prediction rates across racial groups while maintaining an AUC of 0.792 for predicting graft failure.

Li *et al.* further expanded on fairness with a transformer-based architecture that addressed multiple dimensions of bias<sup>[10]</sup>. Their innovative multi-task learning strategy used dynamic reweighting mechanisms to balance performance across different post-transplant risk predictions. This model incorporated fairness metrics, such as demographic parity (ensuring equal prediction rates across demographic groups) and equalized odds (ensuring similar true positive and false positive rates across groups). Their approach achieved substantial reductions in fairness disparities - up to 97% across gender groups and 94% across age groups - while maintaining high predictive accuracy across all tasks.

To ensure fairness, AI studies in LT should, at a minimum, conduct comprehensive subgroup analyses across key demographic factors like age, gender, and race/ethnicity, as demonstrated by Zaver *et al.*<sup>[70]</sup>. Such analyses are essential for identifying potential disparities in model performance and ensuring transparency in reporting how AI models behave across diverse patient populations. Addressing these issues is a crucial step toward the ethical and equitable translation of AI tools into clinical practice.

## INTEGRATING AI INTO CLINICAL WORKFLOW

The successful implementation of AI systems in LT clinical practice (deployment) requires a well-planned deployment strategy and robust integration workflows [Figure 1]. Despite a growing body of research on AI applications in LT, there remains a significant gap in the literature addressing practical implementation strategies. Moreover, there are currently no FDA-approved AI algorithms specifically tailored for LT, highlighting the early stage of clinical AI adoption in this field. Clinical implementation typically follows two primary deployment pathways, as shown in Figure 1: external deployment systems and integration into EHRs. Epic, the leading EHR platform in the United States, recently introduced the EPIC Cognitive Computing platform that facilitates AI integration into EPIC EHR. Epic Nebula supports embedding external AI models.

Early demonstration of clinical impact in live settings is crucial. A practical method is a silent-mode deployment in a pilot setting within the EHR, enabling clinicians to retrospectively assess AI outputs without influencing immediate decisions. Afshar *et al.* described an example of the clinical utilization of AI models by deploying a real-time AI-based clinical decision support tool for opioid misuse screener in a non-transplant setting<sup>[79]</sup>.

Integration of AI tools into clinical workflows presents unique opportunities for perioperative decision support. Pilot studies of such operating room implementations will be essential to demonstrate both technical feasibility and clinical impact before widespread adoption.

Performance monitoring is a critical aspect of AI deployment and involves two key processes: periodic performance assessment and bias evaluation to ensure fairness. Regular monitoring is essential for detecting declines in model performance or emerging biases that could negatively affect patient care [Figure 1]. This step is crucial for maintaining the clinical utility and fairness of AI systems across diverse patient populations. The continuous improvement phase involves resolving performance issues identified during monitoring and adapting to model drift caused by new treatments or evolving clinical policies<sup>[80]</sup>. This iterative process is particularly significant in LT, where protocols and patient demographics change over time.

Data harmonization also poses a significant challenge in AI implementation. Variations in data availability, granularity, and variable standardization across transplant registries limit the generalizability of AI models. For instance, harmonizing variables across datasets has been shown to reduce model performance, with

AUROC scores decreasing from 0.74 to 0.68 for Canadian data and from 0.71 to 0.66 for US data<sup>[55]</sup>. Addressing this challenge is critical for creating AI models that are robust across different healthcare systems and regions.

AI's promise to enhance LT decision making hinges on meeting regulatory requirements. In the United States, the FDA oversees AI under "Software as a Medical Device" (SaMD), ensuring safety, efficacy, and transparent reporting. Similarly, in Europe, CE marking and compliance with the EU Medical Device Regulation (EU MDR) are required. These regulations emphasize rigorous validation, real-world performance data, and ongoing post-market surveillance.

Cost is also one of the most immediate barriers. Health systems must budget for graphics processing unit (GPU) servers or high-availability cloud inference, annual software maintenance fees, and EHR integration licenses. However, they cannot currently recoup these expenses because no current procedural terminology (CPT) or diagnosis-related group (DRG) codes reimburse AI-assisted decision support. Therefore, formal cost-utility analyses and early conversations with payers will be essential before LT centers commit capital.

Ultimately, the success of AI integration in LT relies on establishing a cyclical framework of monitoring, evaluation, and improvement while ensuring compatibility with existing clinical workflows. By addressing challenges such as performance monitoring, fairness evaluation, and data harmonization, this framework provides the foundation for leveraging AI's full potential to transform transplant care delivery.

## FUTURE DIRECTIONS FOR AI IN LT

The future of AI in LT is poised for significant advancements, particularly with the increasing availability of large datasets. Transformer-based models, known for their scalability and ability to process complex data, hold immense potential in this field. Their performance continues to improve as the size of training datasets expands, a principle demonstrated by widely used models like ChatGPT<sup>[81]</sup>. Large language models, such as ChatGPT, can assist clinicians by drafting clinic letters, generating empathetic responses to patient messages, and creating discharge summaries - all while maintaining clinical accuracy<sup>[82,83]</sup>. In research, Generative AI (GenAI) offers the potential to streamline study design, simplify protocol development, and address ethical considerations. Advanced techniques like generative adversarial networks (GANs) and variational autoencoders show promise in drug development, paving the way for personalized immunosuppressive therapies. Moreover, GenAI can enhance education by producing customized learning materials for diverse audiences, including physicians, nurses, and transplant coordinators.

Foundation models, such as Med-BERT<sup>[84]</sup>, demonstrate the potential of large-scale pretraining on EHR data. These models undergo a two-step process: pretraining on large, diverse datasets followed by fine-tuning on transplant-specific data. This approach enables them to capture intricate medical patterns, enhancing outcome predictions. Such models provide a crucial advantage in transplant medicine, where labeled data for rare conditions is often limited.

Advancements in self-supervised and unsupervised learning are unlocking new possibilities for phenotype discovery<sup>[85]</sup>. These approaches analyze clinical, laboratory, and imaging data comprehensively, bypassing the need for predefined categorizations. By identifying complex and multifaceted phenotypes, they offer a data-driven, unbiased perspective that reduces variability inherent in traditional, human-dependent definitions. This innovation provides profound insights into the complexities of transplant medicine.

## PROMOTING ACCESS WITH AI IN LT

In the LT field, when researchers discuss improving allocation through AI, they typically refer to enhancing donor-recipient matching by incorporating both donor and recipient variables to better predict outcomes. A recent systematic review by Pruinelli *et al.* specifically examined AI applications in liver transplant allocation<sup>[3]</sup>. Notably, they found that most models addressed the principle of utility through post-transplant outcomes prediction, with very few studies attempting to improve MELD for waitlist mortality prediction (urgency), and none successfully developing a transplant-related benefit model. This gap aligns with our broader findings across the entire transplant continuum.

AI in LT holds the potential to identify patients at risk of poor outcomes, enabling targeted interventions and optimizing treatment strategies. While AI may inform decisions in transplant prioritization, it must not serve as the sole basis for withdrawing patients from consideration, especially given current limitations. It is crucial to ensure that AI algorithms do not exacerbate existing vulnerabilities or create new barriers to transplant access<sup>[86]</sup>. Prior to AI, efficiency-based algorithms have raised concerns that prioritizing metrics like “the greatest increase in quality-adjusted life expectancy” are systematically disadvantageous to already marginalized groups due to their higher predicted risks of graft failure<sup>[87]</sup>. AI models must, therefore, be designed with a dual focus on utility and equity, incorporating safeguards and proactive measures to address disparities.

High-quality and diverse datasets are essential for training robust AI models; however, data scarcity remains a significant challenge. Many AI studies (62% of included articles) rely on single-center retrospective data, which leads to overfitting to specific institutional practices and underrepresentation of diverse populations. Additionally, challenges such as data heterogeneity, imbalance, and missing values in transplant datasets undermine model generalizability. The interpretability of complex models, such as neural networks and ensemble methods, also presents a major obstacle, despite advancements like SHAP and saliency maps. Furthermore, algorithmic fairness remains a critical concern, with limited attention given to addressing biases related to gender, race, and socioeconomic factors. Such biases can perpetuate disparities in organ allocation and patient outcomes. The lack of multicenter validations, regulatory pathways, and ethical frameworks for transparency and accountability further raises concerns about the readiness of AI for real-world clinical use. Moreover, resource constraints, including the high cost of development and maintenance, limit the feasibility of deploying AI in resource-limited settings.

Although predictive models can identify high-risk cases requiring additional support, using AI as a gatekeeper to categorically restrict organ access is both ethically problematic and scientifically premature, given the current limitations in data quality, fairness, and model interpretability. The way forward lies in developing AI systems that augment rather than replace human decision making, with robust safeguards to prevent discrimination and ensure equitable access for all populations. By embracing this balanced approach, we can harness AI's potential to advance transplant medicine while upholding the fundamental principles of justice and beneficence in organ allocation.

## CONCLUSION

This narrative review underscores the transformative potential of AI in LT, illustrating its diverse applications across the transplant continuum. From AI in pre-transplant risk assessment, donor liver assessment, and transplant oncology to graft survival prediction, overall survival prediction, immunosuppression management, and post-transplant complications prediction, AI has demonstrated significant promise in enhancing clinical decision making, improving predictive accuracy, and streamlining workflows. Emerging advancements, including transformer models, generative AI, and foundation models,

are poised to further elevate these applications and broaden AI's impact in LT. However, to achieve widespread clinical adoption, several critical barriers must be overcome. Key challenges include ensuring algorithmic fairness, enhancing model interpretability, conducting robust multicenter validation, and integrating AI into existing clinical workflows. Given these current and foreseeable limitations, AI must be employed as a tool to improve patient outcomes and ensure equitable care - not as a gatekeeper to deny organ access. With a balanced approach prioritizing fairness, transparency, and inclusivity, AI can fulfill its potential to transform LT while upholding the ethical principles of transplantation medicine.

## DECLARATIONS

### Authors' contributions

Designed and conceptualized the study: Patel KA, Connor AA, Ghobrial RM

Collected data, created figures, and wrote the primary manuscript draft: Patel K

Contributed to the writing of specific sections and critically reviewed the manuscript: Kodali S, Mobley CM, Victor D, Hobeika MJ, Dib Y, Saharia A, Cheah YL, Simon CJ, Brombosz EW, Moore LW

All authors read and approved the final manuscript.

### Availability of data and materials

Not applicable.

### Financial support and sponsorship

None.

### Conflicts of interest

All authors declared that there are no conflicts of interest.

### Ethical approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Copyright

© The Author(s) 2025.

## REFERENCES

1. Gulla A, Jakiunaite I, Juchneviciute I, Dzemyda G. A narrative review: predicting liver transplant graft survival using artificial intelligence modeling. *Front Transplant.* 2024;3:1378378. DOI PubMed PMC
2. Bhat M, Rabindranath M, Chara BS, Simonetto DA. Artificial intelligence, machine learning, and deep learning in liver transplantation. *J Hepatol.* 2023;78:1216-33. DOI PubMed
3. Pruinelli L, Balakrishnan K, Ma S, et al. Transforming liver transplant allocation with artificial intelligence and machine learning: a systematic review. *BMC Med Inform Decis Mak.* 2025;25:98. DOI PubMed PMC
4. Yu YD, Lee KS, Man Kim J, et al; Korean Organ Transplantation Registry Study Group. Artificial intelligence for predicting survival following deceased donor liver transplantation: retrospective multi-center study. *Int J Surg.* 2022;105:106838. DOI PubMed
5. Ge J, Digitale JC, Fenton C, et al. Predicting post-liver transplant outcomes in patients with acute-on-chronic liver failure using expert-augmented machine learning. *Am J Transplant.* 2023;23:1908-21. DOI
6. He ZL, Zhou JB, Liu ZK, et al. Application of machine learning models for predicting acute kidney injury following donation after cardiac death liver transplantation. *Hepatobiliary Pancreat Dis Int.* 2021;20:222-31. DOI PubMed
7. Jain V, Bansal A, Radakovich N, et al. Machine learning models to predict major adverse cardiovascular events after orthotopic liver transplantation: a cohort study. *J Cardiothorac Vasc Anesth.* 2021;35:2063-9. DOI PubMed
8. Yanagawa R, Iwadoh K, Akabane M, et al. LightGBM outperforms other machine learning techniques in predicting graft failure after liver transplantation: creation of a predictive model through large-scale analysis. *Clin Transplant.* 2024;38:e15316. DOI PubMed

9. Yao W, Bai J, Liao W, Chen Y, Liu M, Xie Y. From CNN to transformer: a review of medical image segmentation models. *J Imaging Inform Med*. 2024;37:1529-47. DOI PubMed PMC
10. Li C, Jiang X, Zhang K. A transformer-based deep learning approach for fairly predicting post-liver transplant risk factors. *J Biomed Inform*. 2024;149:104545. DOI PubMed PMC
11. Nitski O, Azhie A, Qazi-Arisar FA, et al. Long-term mortality risk stratification of liver transplant recipients: real-time application of deep learning algorithms on longitudinal data. *Lancet Digit Health*. 2021;3:e295-305. DOI PubMed
12. Papanastasiou G, Dikaos N, Huang J, Wang C, Yang G. Is attention all you need in medical image analysis? A review. *IEEE J Biomed Health Inform*. 2024;28:1398-411. DOI PubMed
13. Wada N, Fujita N, Ishimatsu K, et al. A novel fast kilovoltage switching dual-energy computed tomography technique with deep learning: utility for non-invasive assessments of liver fibrosis. *Eur J Radiol*. 2022;155:110461. DOI PubMed
14. Yu H, Sharifai N, Jiang K, et al. Artificial intelligence based liver portal tract region identification and quantification with transplant biopsy whole-slide images. *Comput Biol Med*. 2022;150:106089. DOI PubMed
15. Ahn JC, Attia ZI, Rattan P, et al. Development of the AI-cirrhosis-ECG score: an electrocardiogram-based deep learning model in cirrhosis. *Am J Gastroenterol*. 2022;117:424-32. DOI PubMed PMC
16. Mazumder NR, Enchakalody B, Zhang P, Su GL. Using artificial intelligence to predict cirrhosis from computed tomography scans. *Clin Transl Gastroenterol*. 2023;14:e00616. DOI PubMed PMC
17. Azhie A, Sharma D, Sheth P, et al. A deep learning framework for personalised dynamic diagnosis of graft fibrosis after liver transplantation: a retrospective, single Canadian centre, longitudinal study. *Lancet Digit Health*. 2023;5:e458-66. DOI PubMed
18. Qazi Arisar FA, Salinas-Miranda E, Ale Ali H, et al. Development of a radiomics-based model to predict graft fibrosis in liver transplant recipients: a pilot study. *Transpl Int*. 2023;36:11149. DOI PubMed PMC
19. Gerussi A, Verda D, Bernasconi DP, et al. Machine learning in primary biliary cholangitis: a novel approach for risk stratification. *Liver Int*. 2022;42:615-27. DOI PubMed
20. Umbaugh DS, Nguyen NT, Curry SC, et al; Acute Liver Failure Study Group. The chemokine CXCL14 is a novel early prognostic biomarker for poor outcome in acetaminophen-induced acute liver failure. *Hepatology*. 2024;79:1352-64. DOI PubMed PMC
21. Schuessler M, Saner F, Al-Rashid F, Schlosser T. Diagnostic accuracy of coronary computed tomography angiography-derived fractional flow reserve (CT-FFR) in patients before liver transplantation using CT-FFR machine learning algorithm. *Eur Radiol*. 2022;32:8761-8. DOI PubMed PMC
22. Ahmed O, Doyle MBM. Liver transplantation: expanding the donor and recipient pool. *Chin Clin Oncol*. 2021;10:6. DOI PubMed
23. Silva AC, Nogueira P, Machado MV. Hepatic steatosis after liver transplantation: a systematic review and meta-analysis. *Liver Transpl*. 2023;29:431-48. DOI PubMed
24. Pérez-Sanz F, Riquelme-Pérez M, Martínez-Barba E, et al. Efficiency of machine learning algorithms for the determination of macrovesicular steatosis in frozen sections stained with sudan to evaluate the quality of the graft in liver transplantation. *Sensors*. 2021;21:1993. DOI PubMed PMC
25. Tang H, Jiao J, Lin JD, Zhang X, Sun N. Detection of large-droplet macrovesicular steatosis in donor livers based on segment-anything model. *Lab Invest*. 2024;104:100288. DOI PubMed
26. Gambella A, Salvi M, Molinaro L, et al. Improved assessment of donor liver steatosis using Banff consensus recommendations and deep learning algorithms. *J Hepatol*. 2024;80:495-504. DOI PubMed
27. Frey KL, McLeod MC, Cannon RM, et al. Non-invasive evaluation of hepatic macrosteatosis in deceased donors. *Am J Surg*. 2023;226:692-6. DOI PubMed
28. Cherchi V, Mea VD, Terrosu G, et al. Assessment of hepatic steatosis based on needle biopsy images from deceased donor livers. *Clin Transplant*. 2022;36:e14557. DOI PubMed
29. Lim J, Han S, Lee D, et al. Identification of hepatic steatosis in living liver donors by machine learning models. *Hepatol Commun*. 2022;6:1689-98. DOI PubMed PMC
30. Jeong JG, Choi S, Kim YJ, Lee WS, Kim KG. Deep 3D attention CLSTM U-Net based automated liver segmentation and volumetry for the liver transplantation in abdominal CT volumes. *Sci Rep*. 2022;12:6370. DOI PubMed PMC
31. Yang X, Park S, Lee S, et al. Estimation of right lobe graft weight for living donor liver transplantation using deep learning-based fully automatic computed tomographic volumetry. *Sci Rep*. 2023;13:17746. DOI PubMed PMC
32. Giglio MC, Zanfardino M, Franzese M, et al. Machine learning improves the accuracy of graft weight prediction in living donor liver transplantation. *Liver Transpl*. 2023;29:172-83. DOI PubMed
33. Kazami Y, Kaneko J, Keshwani D, et al. Two-step artificial intelligence algorithm for liver segmentation automates anatomic virtual hepatectomy. *J Hepatobiliary Pancreat Sci*. 2023;30:1205-17. DOI PubMed
34. Oh N, Kim B, Kim T, Rhu J, Kim J, Choi GS. Real-time segmentation of biliary structure in pure laparoscopic donor hepatectomy. *Sci Rep*. 2024;14:22508. DOI PubMed PMC
35. Bruix J, Reig M, Sherman M. Evidence-based diagnosis, staging, and treatment of patients with hepatocellular carcinoma. *Gastroenterology*. 2016;150:835-53. DOI PubMed
36. Huang H, Xie Y, Wang G, Zhang L, Zhou W. DLNLF-net: denoised local and non-local deep features fusion network for malignancy characterization of hepatocellular carcinoma. *Comput Methods Programs Biomed*. 2022;227:107201. DOI PubMed
37. Kwong A, Hameed B, Syed S, et al. Machine learning to predict waitlist dropout among liver transplant candidates with hepatocellular carcinoma. *Cancer Med*. 2022;11:1535-41. DOI PubMed PMC

38. He T, Fong JN, Moore LW, et al. An imageomics and multi-network based deep learning model for risk assessment of liver transplantation for hepatocellular cancer. *Comput Med Imaging Graph*. 2021;89:101894. DOI PubMed PMC
39. Liu Z, Liu Y, Zhang W, et al. Deep learning for prediction of hepatocellular carcinoma recurrence after resection or liver transplantation: a discovery and validation study. *Hepatol Int*. 2022;16:577-89. DOI PubMed PMC
40. Ivanics T, Nelson W, Patel MS, et al. The Toronto postliver transplantation hepatocellular carcinoma recurrence calculator: a machine learning approach. *Liver Transpl*. 2022;28:593-602. DOI PubMed
41. Liu S, Nalesnik MA, Singhi A, et al. Transcriptome and exome analyses of hepatocellular carcinoma reveal patterns to predict cancer recurrence in liver transplant patients. *Hepatol Commun*. 2022;6:710-27. DOI PubMed PMC
42. Tran BV, Moris D, Markovic D, et al. Development and validation of a REcurrent Liver eAnceR Prediction ScorE (RELAPSE) following liver transplantation in patients with hepatocellular carcinoma: Analysis of the US Multicenter HCC Transplant Consortium. *Liver Transpl*. 2023;29:683-97. DOI PubMed
43. Qu WF, Tian MX, Lu HW, et al. Development of a deep pathomics score for predicting hepatocellular carcinoma recurrence after liver transplantation. *Hepatol Int*. 2023;17:927-41. DOI PubMed PMC
44. To J, Ghosh S, Zhao X, et al. Deep learning-based pathway-centric approach to characterize recurrent hepatocellular carcinoma after liver transplantation. *Hum Genomics*. 2024;18:58. DOI PubMed PMC
45. Iseke S, Zeevi T, Kucukkaya AS, et al. Machine learning models for prediction of posttreatment recurrence in early-stage hepatocellular carcinoma using pretreatment clinical and MRI features: a proof-of-concept study. *AJR Am J Roentgenol*. 2023;220:245-55. DOI PubMed PMC
46. Altaf A, Mustafa A, Dar A, et al. Artificial intelligence-based model for the recurrence of hepatocellular carcinoma after liver transplantation. *Surgery*. 2024;176:1500-6. DOI PubMed
47. Guijo-Rubio D, Briceño J, Gutiérrez PA, Ayllón MD, Ciria R, Hervás-Martínez C. Statistical methods versus machine learning techniques for donor-recipient matching in liver transplantation. *PLoS One*. 2021;16:e0252068. DOI PubMed PMC
48. Cooper JP, Perkins JD, Warner PR, et al. Acute graft-versus-host disease after orthotopic liver transplantation: predicting this rare complication using machine learning. *Liver Transpl*. 2022;28:407-21. DOI PubMed PMC
49. Ding S, Tang R, Zha D, et al. Fairly predicting graft failure in liver transplant for organ assigning. *AMIA Annu Symp Proc*. 2023;2022:415-24. PubMed PMC
50. Lin Y, Huang H, Cao J, et al. An integrated proteomics and metabolomics approach to assess graft quality and predict early allograft dysfunction after liver transplantation: a retrospective cohort study. *Int J Surg*. 2024;110:3480-94. DOI PubMed PMC
51. Bambha K, Kim NJ, Sturdevant M, et al. Maximizing utility of nondirected living liver donor grafts using machine learning. *Front Immunol*. 2023;14:1194338. DOI PubMed PMC
52. Zalba Etayo B, Marín Araiz L, Montes Aranguren M, et al. Graft survival in liver transplantation: an artificial neuronal network assisted analysis of the importance of comorbidities. *Exp Clin Transplant*. 2023;21:338-44. DOI PubMed
53. Yang M, Peng B, Zhuang Q, et al. Models to predict the short-term survival of acute-on-chronic liver failure patients following liver transplantation. *BMC Gastroenterol*. 2022;22:80. DOI PubMed PMC
54. Yasodhara A, Dong V, Azhie A, Goldenberg A, Bhat M. Identifying modifiable predictors of long-term survival in liver transplant recipients with diabetes mellitus using machine learning. *Liver Transpl*. 2021;27:536-47. DOI PubMed PMC
55. Ivanics T, So D, Claasen MPAW, et al. Machine learning-based mortality prediction models using national liver transplantation registries are feasible but have limited utility across countries. *Am J Transplant*. 2023;23:64-71. DOI PubMed
56. Park SJ, Yoon JH, Joo I, Lee JM. Newly developed sarcopenia after liver transplantation, determined by a fully automated 3D muscle volume estimation on abdominal CT, can predict post-transplant diabetes mellitus and poor survival outcomes. *Cancer Imaging*. 2023;23:73. DOI
57. Liu Z, Wu Y, Khan AA, et al. Deep learning-based radiomics allows for a more accurate assessment of sarcopenia as a prognostic factor in hepatocellular carcinoma. *J Zhejiang Univ Sci B*. 2024;25:83-90. DOI PubMed PMC
58. Fonseca ALF, Santos BC, Anastácio LR, et al. Global Leadership Initiative on Malnutrition criteria for the diagnosis of malnutrition and prediction of mortality in patients awaiting liver transplant: a validation study. *Nutrition*. 2023;114:112093. DOI PubMed
59. Ding S, Tan Q, Chang CY, et al. Multi-task learning for post-transplant cause of death analysis: a case study on liver transplant. *AMIA Annu Symp Proc*. 2024;2023:913-22. PubMed PMC
60. Rogers MP, Janjua HM, Read M, et al. Recipient survival after orthotopic liver transplantation: interpretable machine learning survival tree algorithm for patient-specific outcomes. *J Am Coll Surg*. 2023;236:563-72. DOI PubMed
61. Ponthier L, Marquet P, Moes DJAR, et al. Application of machine learning to predict tacrolimus exposure in liver and kidney transplant patients given the MeltDose formulation. *Eur J Clin Pharmacol*. 2023;79:311-9. DOI PubMed
62. Li ZR, Li RD, Niu WJ, et al. Population pharmacokinetic modeling combined with machine learning approach improved tacrolimus trough concentration prediction in Chinese adult liver transplant recipients. *J Clin Pharmacol*. 2023;63:314-25. DOI PubMed
63. Du Y, Zhang Y, Yang Z, et al. Artificial neural network analysis of determinants of tacrolimus pharmacokinetics in liver transplant recipients. *Ann Pharmacother*. 2024;58:469-79. DOI PubMed
64. Yoon SB, Lee JM, Jung CW, et al. Machine-learning model to predict the tacrolimus concentration and suggest optimal dose in liver transplantation recipients: a multicenter retrospective cohort study. *Sci Rep*. 2024;14:19996. DOI PubMed PMC
65. Zhang Y, Yang D, Liu Z, et al. An explainable supervised machine learning predictor of acute kidney injury after adult deceased donor liver transplantation. *J Transl Med*. 2021;19:321. DOI PubMed PMC

66. Chen C, Yang D, Gao S, et al. Development and performance assessment of novel machine learning models to predict pneumonia after liver transplantation. *Respir Res*. 2021;22:94. DOI PubMed PMC
67. Freire MP, Rinaldi M, Terrabuio DRB, et al. Prediction models for carbapenem-resistant Enterobacterales carriage at liver transplantation: a multicenter retrospective study. *Transpl Infect Dis*. 2022;24:e13920. DOI PubMed
68. Chen C, Chen B, Yang J, et al. Development and validation of a practical machine learning model to predict sepsis after liver transplantation. *Ann Med*. 2023;55:624-33. DOI PubMed PMC
69. Jang HY, Han SB, Jeong JH, et al. Prognostic value of mitral annular calcification in liver transplant patients: implication in posttransplant outcomes. *Transplantation*. 2024;108:1954-61. DOI PubMed
70. Zaver HB, Mzaik O, Thomas J, et al. Utility of an artificial intelligence enabled electrocardiogram for risk assessment in liver transplant candidates. *Dig Dis Sci*. 2023;68:2379-88. DOI PubMed PMC
71. Fodor M, Zelger P, Pallua JD, et al. Prediction of biliary complications after human liver transplantation using hyperspectral imaging and convolutional neural networks: a proof-of-concept study. *Transplantation*. 2024;108:506-15. DOI PubMed
72. Cheng Y, Zhang XD, Chen C, et al. Dynamic evolution of brain structural patterns in liver transplantation recipients: a longitudinal study based on 3D convolutional neuronal network model. *Eur Radiol*. 2023;33:6134-44. DOI PubMed
73. Lee BP, Roth N, Rao P, et al. Artificial intelligence to identify harmful alcohol use after early liver transplant for alcohol-associated hepatitis. *Am J Transplant*. 2022;22:1834-41. DOI PubMed PMC
74. Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell*. 2020;2:56-67. DOI PubMed PMC
75. Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?": explaining the predictions of any classifier. *arXiv* 2016; arXiv:1602.04938. Available from: <https://doi.org/10.48550/arXiv.1602.04938>. [Last accessed on 15 May 2025]
76. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMJ*. 2015;350:g7594. DOI PubMed
77. Wood NL, VanDerwerken D, Segev DL, Gentry SE. Correcting the sex disparity in MELD-Na. *Am J Transplant*. 2021;21:3296-304. DOI PubMed PMC
78. Strauss AT, Sidoti CN, Sung HC, et al. Artificial intelligence-based clinical decision support for liver transplant evaluation and considerations about fairness: a qualitative study. *Hepatol Commun*. 2023;7:e0239. DOI PubMed PMC
79. Afshar M, Adelaine S, Resnik F, et al. Deployment of real-time natural language processing and deep learning clinical decision support in the electronic health record: pipeline implementation for an opioid misuse screener in hospitalized adults. *JMIR Med Inform*. 2023;11:e44977. DOI PubMed PMC
80. Sahiner B, Chen W, Samala RK, Petrick N. Data drift in medical machine learning: implications and potential remedies. *Br J Radiol*. 2023;96:20220878. DOI PubMed PMC
81. Deeb M, Gangadhar A, Rabindranath M, et al. The emerging role of generative artificial intelligence in transplant medicine. *Am J Transplant*. 2024;24:1724-30. DOI PubMed
82. Ayers JW, Poliak A, Dredze M, et al. Comparing physician and artificial intelligence Chatbot responses to patient questions posted to a public social media forum. *JAMA Intern Med*. 2023;183:589-96. DOI PubMed PMC
83. Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI Chatbot for medicine. *N Engl J Med*. 2023;388:1233-9. DOI PubMed
84. Rasmy L, Xiang Y, Xie Z, Tao C, Zhi D. Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction. *NPJ Digit Med*. 2021;4:86. DOI PubMed PMC
85. Patel K, Xie Z, Yuan H, et al. Unsupervised deep representation learning enables phenotype discovery for genetic association studies of brain imaging. *Commun Biol*. 2024;7:414. DOI PubMed PMC
86. Lebret A. Allocating organs through algorithms and equitable access to transplantation-a European human rights law approach. *J Law Biosci*. 2023;10:lsad004. DOI PubMed PMC
87. Zenios SA, Wein LM, Chertow GM. Evidence-based organ allocation. *Am J Med*. 1999;107:52-61. DOI PubMed