88891_Auto_Edited.docx

WORD COUNT 7901
TIME SUBMITTED 09-NOV-2023 12:02PM
PAPER ID 104097520

Name of Journal: *World Journal of Transplantation*
Manuscript NO: 88891
Manuscript Type: SYSTEMATIC REVIEWS

Use of Machine Learning Models for the Prognostication of Liver Transplantation: A systematic review

Machine Learning Models and Prognostication of Liver Transplantation

Gidion Chongo, Jonathan Soldera

Abstract

BACKGROUND

Background: Liver transplantation is a life-saving intervention for patients with end-stage liver disease. However, the equitable allocation of scarce donor organs remains a formidable challenge. Prognostic tools are pivotal in identifying the most suitable transplant candidates. Traditionally, scoring systems like the Model for End-Stage Liver Disease (MELD) have been instrumental in this process. Nevertheless, the landscape of prognostication is undergoing a transformation with the integration of machine learning (ML) and artificial intelligence (AI) models.

AIM

Aim: This systematic review aims to assess the utility of ML models in prognostication for liver transplantation, comparing their performance and reliability to established traditional scoring systems.

METHODS

Methods: Following the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines, we conducted a thorough and standardized literature

search using the PubMed/MEDLINE database. Our search imposed no restrictions on publication year, age, or gender. Exclusion criteria encompassed non-English studies, review articles, case reports, conference papers, studies with missing data, or those exhibiting evident methodological flaws.

RESULTS

Results: Our search yielded a total of 64 articles, with 23 meeting the inclusion criteria. Among the selected studies, 60.8% originated from the United States and China combined. Only one pediatric study met the criteria. Notably, 91% of the studies were published within the past five years. ML models consistently demonstrated satisfactory to excellent AUROC values (ranging from 0.6 to 1) across all studies, surpassing the performance of traditional scoring systems. Random Forest (RF) exhibited superior predictive capabilities for 90-day mortality following liver transplantation, sepsis, and acute kidney injury (AKI). In contrast, gradient boosting excelled in predicting the risk of graft-versus-host disease (GVHD), pneumonia, and AKI.

CONCLUSION

Conclusion: This study underscores the potential of ML models in guiding decisions related to allograft allocation and liver transplantation, marking a significant evolution in the field of prognostication.

Key Words: Keywords: Liver Transplantation; Machine Learning Models; Prognostication; Allograft Allocation; Artificial Intelligence.

Chongo G, Soldera J. Use of Machine Learning Models for the Prognostication of Liver Transplantation: A systematic review. *World J Transplant* 2023; In press

Core Tip: Core Tip: This systematic review highlights the promising role of machine learning (ML) models in improving prognostication for liver transplantation. ML

models consistently outperformed traditional scoring systems, demonstrating excellent predictive capabilities for various post-transplant complications, including mortality, sepsis, and acute kidney injury. The findings underscore the potential of ML in enhancing decision-making related to organ allocation and liver transplantation, representing a substantial advancement in prognostication methods.

INTRODUCTION

INTRODUCTION

Liver transplantation (LT) has long been a transformative intervention for individuals afflicted with acute and chronic-end-stage liver ailments. In addition to restoring patients' health, liver transplantation can enhance their overall well-being and potentially extend their lifespan by up to 15 years [1]. This treatment approach is firmly established as a last resort when alternative methods and therapies have proven ineffective. According to the Scientific Registry of Transplant Recipients in the United States, the survival rates for patients after deceased donor liver transplantation are commendable, standing at approximately 90% at one year and 77% at five years post-LT [2]. Nevertheless, the field of liver transplantation confronts a range of challenges, encompassing candidate selection, organ allocation, and a scarcity of donor organs.

The persistent scarcity of donor organs has emerged as a critical and ongoing concern. While living donation has bolstered liver transplant numbers in some regions, in others, the field has stagnated. Consequently, there has been a concerted effort over the past decade to augment the pool of deceased donors. This endeavor has led to increased utilization of liver allografts obtained after cardiac death (DCD), as well as those from marginal and extended donor criteria [3]. Despite these improvements, a notable number of DCD livers remain unused due to suboptimal allograft function and unacceptable donor parameters. This predicament has given rise to the concept of mechanical perfusion for solid organ transplantation, aiming to expand the available organ pool,

particularly for liver allografts, further underscoring the significant scarcity of this vital resource for transplantation [4].

A recent study emphasized the multifaceted challenges inherent to liver transplantation. In 2017, the United States recorded a waiting list of 14,360 candidates eagerly awaiting liver transplantation [5]. Furthermore, the study reported an average hospital expenditure exceeding $490,000 per patient associated with liver transplantation in 2011 [5]. Evidently, there is an escalating demand for a more efficient system of liver organ allocation to optimize outcomes within a society grappling with diminishing liver organ donations and escalating expenditures linked to the care of end-stage liver disease patients.

The allocation of liver allografts to patients in need has relied on various scoring tools. Initially, the Child-Pugh-Turcotte (CTP) score served this purpose, but the Model for End-stage Liver Disease (MELD) has now become the preferred score for organ allocation. Additionally, several other scoring systems, such as Survival Outcomes Following Liver Transplantation (SOFT), Balance of Risk (BAR), Donor Risk Index (DRI), ABIC, CLIF-C OFs, CLIF-C ACLFs, and CLIF-SOFA, have been employed in this context.

The CTP score, initially validated for predicting postoperative mortality in cirrhotic patients, incorporates clinical and biochemical data, including serum albumin, serum bilirubin, international normalized ratio (INR) or prothrombin time (PT), ascites, and encephalopathy, to assess the prognosis of end-stage liver disease. The total CP score is calculated by assigning points to each variable, with a maximum score of 15 points. CP class A corresponds to a score of 5-6 points, with a 10% mortality rate. CP class B corresponds to a score of 7-9 points, with a 30% mortality rate, while CP class C represents a score of 10-15 points, associated with a poorer prognosis, including a 50% mortality rate at one-to-five years and sometimes as high as 70-80% [6] [7] [8].

However, the use of CTP for liver transplant allocation had significant limitations. It relied on subjective assessments of ascites and encephalopathy, lacked an evaluation of renal function, and had a limited scoring range, making it challenging to differentiate

patients based on disease severity. This limitation was evident when patients with different INR and bilirubin levels were assigned the same CTP score, potentially leading to misleading prioritization [9]. Other drawbacks of the CTP score include the empirical selection of variables and the interdependence of some variables, such as coagulation and albumin, which could result in an imbalance in their influence within the score.

The CTP score's arbitrary cutoffs for quantitative variables lack evidence of optimality in defining hepatic changes and mortality risk, hindering its reliability in predicting prognosis in liver cirrhosis and post-liver transplantation [10]. Conversely, MELD score, originally designed for predicting survival after trans-jugular intrahepatic Porto-systemic shunt (TIPS) procedures, has been extended to assess prognosis in liver cirrhosis and serves as a tool for liver organ allocation [11]. MELD score's has a good reliability in predicting 1-year and 5-year survival across diverse liver diseases, including alcoholic cirrhosis and hepatitis [12]. Additionally, MELD score has prognostic value in conditions like spontaneous bacterial peritonitis, variceal bleeding, and hepatorenal syndrome (HRS) [13]. In cases of variceal bleeding, the MELD score's predictive ability was comparable to the CTP score. Concerning HRS, a high MELD score (>20) has been linked to a median survival of just 1 mo for type 1 HRS, while type 2 HRS patients' survival correlated with their MELD score, with a median survival of 3 mo for MELD >20 and 11 mo for MELD <20 [14]. To enhance its predictive power, the MELD score has evolved into multiple versions, including MELD sodium (MELD NA) and Delta MELD (D-MELD).

MELD NA, developed due to the observation of dilutional hyponatremia in cirrhotic patients, stems from systemic arterial vasodilation-induced antidiuretic hormone release, which was linked to portal hypertension severity [15]. Hyponatremia indirectly contributes to portal hypertension, leading to complications like ascites, hepatorenal syndrome, and liver-related mortality [16]. Neurologic dysfunction, refractory ascites, hepatorenal syndrome, and liver disease-related death are also associated with hyponatremia [17]. Numerous studies affirm hyponatremia as an independent predictor

of early mortality, with the most pronounced impact between sodium concentrations of 120 to 135 mEq/L. A 1 mEq/L decrease corresponds to a 12% reduction in 3-month survival probability. Adding sodium to the MELD score enhances its predictive accuracy, especially for lower MELD scores. However, this addition doesn't significantly improve survival prediction at 3 and 12 mo and has its limitations due to fluctuating serum sodium levels influenced by various factors [18,19].

The D-MELD was introduced to address the limitation of a single MELD score at a specific time. While it is useful in predicting survival in cirrhotic patients awaiting transplantation, conflicting evidence exists. The potential bias in frequent laboratory testing for acutely worsening patients also complicates its use [20,21]. In summary, all versions of the MELD score have limitations, including susceptibility to therapeutic interventions, empirical variable selection, limited predictive ability for post-transplant mortality, and the need for on-site computation [10].

To improve the prediction of post-liver transplant mortality, various prediction tools have been explored, including the Donor Risk Index (DRI), Enhanced DRI (ET-DRI), SOFT, p-SOFT, BAR, ABIC, CLIF C OFS, CLIF-C ACLFS, and the CLIF-SOFA. The DRI, predating the MELD score, was initially considered as an independent predictor of allograft failure across different MELD categories. However, numerous studies have revealed its limited association with outcomes [22]. The DRI's limitations include its validation in the pre-MELD era, the absence of recipient-related risk factors as the fact that is impractical for predicting morbidity and graft failure due to its poor predictive ability, inclusion of irrelevant factors (e.g., ethnicity), and omission of relevant factors [23].

The Eurotransplant-DRI (ET-DRI) replaces ethnicity and height risk factors with parameters like the latest GGT and rescue offer in the Eurotransplant context. Although it has been shown to be potentially useful for liver allocation, studies have consistently shown its limited predictive ability for early post-transplant outcomes [22-26]. Overall, the ET-DRI is consistently considered an unreliable tool for predicting morbidity and mortality after liver transplantation.

Various prediction tools have been explored to enhance post-liver transplant prognostication. The Survival Outcomes following liver transplant (SOFT) scorehas been tested for predicting 90-day post-transplant mortality [27], [22]. A derivative of SOFT, the pre-allocation SOFT (p-SOFT) score, exhibited promising predictive accuracy [22]. However, the complexity of these scores, which involve multiple subjective and semi-quantitative variables, hampers their prompt clinical assessment and decision-making. Furthermore, their predictive ability for major morbidity at 3 mo appears limited [22] [28].

The Balance of risk (BAR) score offers promise by evaluating both recipient and donor factors for severe complications and 90-day mortality [22, 28]. This tool has shown robustness in various patient populations, including pediatric, adolescent, and living donor liver transplant recipients [29] [30]. However, in specific patient subgroups, BAR's accuracy in assessing short-term outcomes, including major complications, 90-day mortality, and ICU and hospital stay length, may be suboptimal [22].

The Age, Bilirubin, INR, and Creatinine (ABIC) score aim to predict outcomes in patients with alcoholic hepatitis. While it has shown potential, its validation has been inconsistent, and it may not be widely applicable. Additionally, it primarily assesses the risk of wait-time mortality, making it unsuitable for post-liver transplant mortality assessment [31], [32].

The Chronic liver failure-SOFA (CLIF-SOFA) score, a modified version of the Sequential Organ Failure Assessment score (SOFA), is tailored for end-stage liver disease patients. This adaptation replaces platelet count and Glasgow coma scale with INR and hepatic encephalopathy, respectively. Additionally, it incorporates terlipressin and renal replacement therapy into cardiovascular and renal parameters, respectively, and includes SpO2/FiO2 as an alternative respiratory parameter for patients without an arterial line [33].

In a study published in 2014, the CLIF-SOFA score proved to be a significant predictor of 1-year post-LT mortality, surpassing the SOFA score in discriminatory power on several post-transplant days [34]. CLIF-SOFA score exhibited greater numerical

differences between 1-year survivor and non-survivor groups, especially post-LT. Furthermore, CLIF-SOFA score trends reflected patients' responses to therapeutic strategies, with a CLIF-SOFA score >8 on post-transplant day 7 indicating delayed recovery from multiple organ dysfunction, associated with higher acute rejection rates and poorer 1-year survival rates.

The CLIF-Consortium Organ Failure score (CLIF-C OFs), a simplified version of CLIF-SOFA, uses a 3-point range per organ system and performs similarly to CLIF-SOFA, outperforming SOFA [35]. This score has proven to be an excellent prognostic tool for short-term outcomes in liver transplantation. Another variation, the CLIF-Consortium score for Acute on Chronic Liver Failure (CLIF-C ACLFs), designed for acute-on-chronic liver failure patients, includes the CLIF-SOFA score, age, and white-cell count. Jalan *et al* (2014) demonstrated the superiority of the CLIF-ACLF score in terms of performance compared to CLIF-SOFA and CLIF-C OFs scores [35]. However, inferior performance of CLIF-ACLF compared to CLIF-SOFA has been reported [34]. Results of CLIF-SOFA, CLIF-C [36-39] and ACLF classification [40-43] has been conflicting [7]

In response to the limitations of existing prognostic scores, there is a growing interest in harnessing Machine Learning (ML) models and algorithms to enhance the prediction of outcomes in LT. ML models serve as a bridge between organ allocation and achieving optimal results, capitalizing on the increasing use of artificial intelligence in medicine over the past decade (Figure 1). ML algorithms, as illustrated in Figure 2, rely on various types of input data, including structured, semi-structured, and unstructured data. Structured data, characterized by well-defined formats and adherence to specific data models, is organized in a tabular fashion and includes information like names, dates, and addresses. Semi-structured data, found in NoSQL databases, JSON documents, HTML, and XML, possesses organizational properties that enable analysis. On the other hand, unstructured data, comprising text and multimedia materials from sources like emails, sensor data, and web pages, lacks predefined formats, making it more challenging to process and analyze. To extract valuable insights from data for building intelligent applications in specific problem domains, various ML techniques

are applied based on their learning capabilities. Mohammed *et al* categorized machine learning algorithms into four main groups: supervised, unsupervised, semi-supervised, and reinforcement learning [44,45]. Supervised learning involves mapping input to output based on labeled training data, typically used for tasks like classification and regression. Unsupervised learning, on the other hand, analyzes unlabeled datasets without human intervention and is employed for tasks such as clustering and dimensionality reduction, focusing on extracting generative features and identifying meaningful trends.

In the realm of machine learning, several techniques are employed to enhance predictive models for various applications, including liver transplantation prognostication. One such technique is semi-supervised learning, which effectively leverages both labeled and unlabeled data to achieve improved prediction outcomes, especially when labeled data is limited. This approach plays a crucial role in bridging the gap between supervised and unsupervised learning methods, finding utility in domains such as machine translation, data labeling, and text classification [46].

Reinforcement learning, on the other hand, offers a distinct approach by focusing on environment-driven algorithms that enable software agents and machines to autonomously evaluate optimal behavior within specific contexts. This methodology relies on the concept of rewards and penalties, aiming to utilize insights gained from interactions with the environment to maximize rewards or minimize risks. While reinforcement learning possesses significant potential in training AI models, it is better suited for complex scenarios rather than straightforward problems [47].

Within the realm of classification algorithms, several notable methods find application in health-related domains. Logistic Regression (LR) stands as a commonly used technique, relying on logistic functions to estimate probabilities. While LR can excel in linearly separable datasets, it may suffer from overfitting in high-dimensional scenarios. Regularization techniques like L1 and L2 regularization are often employed to mitigate this issue [46].

Support Vector Machine (SVM) is another prominent classification method with applications in health data. SVM operates in high-dimensional spaces by constructing hyperplanes that maximize the margin between data points in different classes. The choice of kernel functions, such as polynomial, linear, radial basis function (RBF), and sigmoid, significantly influences SVM's performance. However, SVM's efficacy can diminish in the presence of noisy datasets and overlapping target classes [46].

Random Forest (RF) offers a distinct ensemble classification technique, widely used in machine learning and data science applications. RF employs parallel ensembling, training multiple decision tree classifiers on different data subsets and combining their outcomes through averaging or majority voting. This approach effectively addresses overfitting concerns and enhances prediction accuracy, making it suitable for both continuous and categorical data in classification and regression problems [40].

Additionally, Adaptive Boosting (AdaBoost) serves as a valuable classification algorithm in the realm of health data. It adopts a sequential ensembling approach to improve the performance of weak classifiers by learning from their errors. By combining multiple underperforming classifiers, AdaBoost creates a robust classifier with high accuracy, boosting the performance of decision trees, base estimators, and binary classification tasks. However, it's essential to note that AdaBoost can be susceptible to overfitting and sensitivity to noisy data and outliers [48].

These various machine learning techniques have been instrumental in addressing complex problems in health-related domains, including liver transplantation prognostication. However, they also come with their own set of challenges, such as overfitting and interpretability issues. Therefore, periodic reviews are crucial to evaluate their performance and reliability compared to traditional scoring methods. This study aims to conduct a systematic review of observational studies, assessing the effectiveness of machine learning models in liver transplantation prognostication and comparing their performance with established scoring systems.

Extreme Gradient Boosting (XGBoost) stands out as a prominent classifier, belonging to the ensemble learning algorithm family, akin to Random Forest. XGBoost represents a

specific variant of gradient boosting that intricately considers detailed approximations when determining the optimal model. It effectively addresses overfitting concerns by minimizing the loss function and employing advanced regularization techniques, including L1 and L2 regularization. These regularization methods are implemented through the computation of second-order gradients of the loss function, resulting in enhanced model generalization and performance [48].

In the domain of machine learning, Artificial Neural Networks (ANN) and deep learning techniques hold significant sway. Deep learning, a subset of ANN-based approaches, encompasses representation learning and comprises multiple layers, including input, hidden, and output layers. These layers collaboratively facilitate learning from data, giving rise to a computational architecture that excels, particularly when dealing with large datasets. Notable deep learning algorithms encompass Multilayer Perceptron (MLP), Long Short-Term Memory Recurrent Neural Network (LSTM-RNN), Convolutional Neural Network (CNN), and ConvNet, among others [49].

Machine learning demonstrates versatility by not only addressing diagnostic challenges but also serving as a valuable tool in prognostic applications. It proves beneficial in disease prediction, data pattern identification, extraction of medical insights, and patient management [50]. Nevertheless, machine learning models are not without their limitations, as highlighted earlier. Concerns encompass overfitting, interference phenomena, where new data may disrupt previous learning, and the black box dilemma, which pertains to the challenge of explaining model results [51].

Within the context of liver transplantation, machine learning models have garnered increasing attention, underscoring the need for periodic assessments of their reliability and performance compared to conventional scoring systems. To this end, this study endeavors to conduct a systematic review of observational studies. The objective is to comprehensively evaluate the evidence concerning the deployment of machine learning models for prognostication in liver transplantation. This evaluation encompasses an assessment of their performance and reliability, juxtaposed with the array of traditional scoring systems currently available.

6

MATERIALS AND METHODS

METHODS

This systematic review adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA-P) guidelines to ensure a standardized approach [52].

Search Strategy:

A comprehensive literature search was conducted using the PubMed/MEDLINE search engine by one researcher. The search strategy included the following terms: ("Machine Learning" OR "Artificial Intelligence") AND ("Liver Transplantation" OR "Allograft liver") AND ("Prognosis" OR "Mortality" OR "Prognostication"). A reference manager tool, Zotero, was utilized for sorting and managing references.

Study Selection:

All observational studies discussing machine learning models and prognosis of liver transplantation, regardless of the year of publication, age, or sex, were included. Studies written in English were considered. Additionally, studies examining machine learning models and the risk of post-transplant complications were included, as these complications often contribute to transplant failure or mortality. Exclusion criteria encompassed non-English papers, review articles, case reports, conference articles, studies with missing data, or studies with evident methodological flaws.

Data Extraction and Synthesis:

The systematic search was conducted by one reviewer, who screened the potential

studies based on their titles and abstracts. Full-text versions of eligible studies were obtained and thoroughly analyzed for content and methodology.

A summary of the included studies was created, providing a narrative overview of each paper's objectives, methods, results, and conclusions. After reviewing the full papers, data on various elements was extracted including; study type, population studied and year of study, purpose of the study, setting of the study, its methods and results, conclusion, limitations and strengths of the study as well as a summary of the study. Additionally, if reported by the studies, a comparison was made between traditional scores and algorithms *vs* machine learning models. This analysis aimed to explore the performance and effectiveness of machine learning approaches in prognosing liver transplantation outcomes.

By systematically extracting relevant information from the selected studies, a comprehensive understanding of the role of artificial intelligence in liver transplantation prognosis was obtained. The data synthesis process involved organizing and presenting the findings in a coherent manner, allowing for a comprehensive evaluation of the current literature in this field.

This approach enabled to examine the various methodologies employed in the studies, identify key trends, and evaluate the potential benefits and limitations of using machine learning models for prognostication in liver transplantation. The synthesized data from the included studies will contribute to providing valuable insights into the current state of research on the role of artificial intelligence in predicting outcomes in liver transplantation.

RESULTS

RESULTS

Using the predetermined search strategy, a total of 64 references were initially identified. Among these, 7 references were excluded as they were conference articles or review papers. Additionally, 1 duplicate article was removed, and 8 articles were

excluded as they were abstracts only and could not be accessed for full-text reading. Subsequently, a thorough evaluation of the remaining 48 articles was conducted through full-text reading and content analysis.

Following the comprehensive assessment, 23 studies met the inclusion criteria and were included in the final analysis. The selection process and reasons for exclusion of certain studies are visually represented in **Figure 3**, which depicts the flowchart illustrating the search strategy employed. **Table 3**, summarizes the findings of every study included.

Quality assessment

The majority of the included studies were considered to be of good quality, despite being observational in nature and not appraised using any specific quality assessment tool. Many of these studies incorporated validation sets in their analyses, which contributes to the robustness of their findings.

Study outcomes:

The studies assessed in this systematic review covered a range of transplantation reasons, including acute-on-chronic liver failure (ACLF) from various causes, primary sclerosing cholangitis (PSC), and hepatocellular carcinoma (HCC). Among the 23 studies analyzed, the highest number (8 studies, accounting for 34.8%) were conducted in America, followed by 6 studies (26%) from China. Additionally, 2 studies (8.7%) were from Korea, while the remaining studies originated from Spain, Australia, Portugal, Taiwan, Iran, and Brazil, each contributing 1 study (4.3%). Furthermore, there was one multinational study involving participants from the USA, Canada, and the UK, which represented 4.3% of the total sample as depicted by **figure 4**.

The studies analyzed in this review spanned from 2014 to 2023. Notably, the highest proportion of studies (26%, 6 studies) were published in 2021, followed by 5 studies (21.7%) from 2022. Studies from 2019, 2020, 2018, and 2023 accounted for 13% (3 studies) each, while 2014 and 2015 each contributed 1 study (4.3%) as shown in **figure 5**.

Regarding the age of participants, one study involved individuals under 18 years old, while the remaining 22 studies focused on adults aged 18.

Primary outcomes and findings

The primary outcomes of interest in the included studies were mortality and the emergence of complications post liver transplant. Most of the studies reported the receiver operating characteristic (ROC) curve and used the area under the ROC curve (AUROC) as a measure of predictive performance. AUROC values were categorized as excellent (0.9-1), very good (0.8-0.9), good (0.7-0.8), satisfactory (0.6-0.7), and unsatisfactory (0.5-0.6) based on previous classification [75].

Across all the studies, machine learning algorithms and models were developed using pre-transplant donor and/or recipient variables. Short-term mortality predictions were typically up to 90 days, while long-term predictions extended up to 5 years. Analysis of AUROC demonstrated that machine learning models consistently yielded satisfactory to excellent results in predicting short and long-term mortality or the risk of complications post liver transplant.

Furthermore, the AUROC analysis revealed that machine learning models outperformed traditional models and scoring systems, including commonly used models such as MELD, D-MELD, SOFT, P-SOFT, BAR, DRI score, ABIC, CLIF-C OFs, CLIF-C ACLFs, and CLIF SOFA. Additionally, machine learning models showed superiority over models based on Cox and logistic regression. Detailed comparisons and findings are presented in **Table 3**.

Sub-analysis:

In terms of predicting 90-day mortality, the Random Forest (RF) model demonstrated the highest area under the curve (AUC) value of 0.940 compared to other machine learning models. Additionally, among the six studies identified in the literature search that discussed the prediction of complications post liver transplant using machine

learning models, an analysis of the AUC values indicated that the 'gradient boosting machine' model performed better than other machine learning models in predicting the risk of graft-versus-host disease (GVHD), pneumonia, and acute kidney injury (AKI). On the other hand, the RF model showed better performance in predicting the risk of sepsis and AKI post liver transplant. Detailed results and comparisons are provided in **Table 3**.

This sub-analysis highlights the specific performance of machine learning models in predicting 90-day mortality and the risk of complications following liver transplantation. The Random Forest model exhibited superior predictive capability for mortality within the 90-day timeframe.

Furthermore, when examining the prediction of post-transplant complications, the 'gradient boosting machine' model demonstrated better performance in predicting GVHD, pneumonia, and AKI, while the RF model showed greater effectiveness in predicting the risk of sepsis and AKI. These findings emphasize the potential of machine learning techniques in enhancing prognostic accuracy and tailoring clinical management strategies in liver transplantation.

DISCUSSION

DISCUSSION

The review highlights a limited number of studies, just 64, that have explored the application of ML models in the context of liver transplantation. This scarcity of research, despite an unrestricted search, indicates a historical lack of emphasis on the potential of ML models in the realm of prognosis and transplant decision-making. Factors contributing to this limited attention include lingering perceptions of ML models as associated with science fiction and concerns regarding potential errors and patient harm. However, it's noteworthy that ML models have advanced in sophistication and have implemented strategies to address challenges like overfitting.

Their effectiveness is contingent upon access to substantial datasets for continuous learning and refinement [76].

In recent years, there has been a notable surge in research at the intersection of ML and liver transplantation, particularly within the last five years. Among the 23 studies reviewed, a substantial majority (91%) were conducted between 2018 and 2023, signifying a burgeoning interest in this field [77]. Additionally, a significant proportion of these studies (61%) originated from the United States and China. A multinational study involving participants from the USA, UK, and Canada stands out, as it evaluated the 90-day predictive capacity of ML models post-liver transplantation across these countries, utilizing transplant registries. Notably, the study revealed that ML model performance varied when applied across countries, indicating limited external validity. Therefore, it is suggested that ML algorithms should be tailored to each country's specific transplant registry data for enhanced reliability. The underrepresentation of other countries in these studies underscores the importance of more diverse ML research to benefit liver transplant patients worldwide.

Crucially, ML methods employed for the allocation of orthotopic liver transplants, whether from living donors, deceased donors, or cadaveric sources, should be rooted in population-specific parameters pertaining to the recipient. This individualized approach is essential to ensure post-transplant longevity and minimize the risk of complications. The utilization of ML models that take into account an individual's unique population parameters or variables to assess the risk of mortality prior to transplantation holds the potential to prevent unnecessary mortality and morbidity associated with high-risk transplantations [78].

Concerning the underlying reasons for transplantation, factors such as Acute-on-Chronic Liver Failure (ACLF), Primary Sclerosing Cholangitis (PSC), and Hepatocellular Carcinoma (HCC) have been prominent considerations. Existing studies have demonstrated the pivotal role of liver transplantation as a life-saving intervention for ACLF patients. ACLF can manifest at any stage of chronic liver disease, leading to a rapid deterioration in liver function and a high mortality rate within a short timeframe

[80], as it is noticeable a high mortality rate for non-transplanted ACLF patients within 28 and 90 days [81] [82].

Liver transplantation is a critical treatment option for various liver-related conditions, including Acute-on-Chronic Liver Failure (ACLF), Primary Sclerosing Cholangitis (PSC), and Hepatocellular Carcinoma (HCC). However, the efficacy of liver transplantation in ACLF patients remains debated, with conflicting findings suggesting no significant survival advantage over non-transplanted patients [83]. Machine Learning (ML) models have the potential to improve the assessment of short-term mortality risk in ACLF patients post-transplantation, thereby aiding in the allocation of liver allografts and potentially enhancing outcomes [79]. It is imperative to expand the scope of research on ML models in liver transplantation to encompass diverse patient populations, thereby increasing the external validity of these models. Customizing ML algorithms to specific transplant registries and incorporating population-specific parameters can enhance the accuracy and effectiveness of prognosis and decision-making in liver transplantation.

Primary Sclerosing Cholangitis (PSC) is a chronic liver disease characterized by progressive bile duct inflammation, cholestasis, and fibrosis. Liver transplantation is the primary treatment for end-stage PSC, yielding generally favorable outcomes, although complications like cholangiocarcinoma, recurrent disease, worsening of inflammatory bowel disease (IBD), and an elevated risk of colonic cancer pose challenges [84]. Cholangiocarcinoma develops in 8-18% of long-standing PSC patients [85], and PSC recurrence post-transplantation is observed in some cases [86]. Increased dysplasia and colon cancer risk are also associated with colitis patients having coexisting PSC [87] [88]. Consequently, accurate evaluation and allocation of liver allografts in PSC patients are critical, with ML algorithms incorporating pertinent variables from PSC patients facilitating informed and precise decision-making [86-89].

Hepatocellular carcinoma (HCC) is a common indication for liver transplantation, ranking fifth among the most prevalent malignancies and being the third leading cause of cancer-related mortality worldwide [90-92]. Liver transplantation offers a promising

therapeutic option for long-term survival in HCC cases by addressing both advanced liver disease and HCC itself [93]. However, the risk of HCC recurrence post-transplantation underscores the necessity for careful patient selection. HCC recurrence occurs most frequently among liver transplant recipients compared to other liver diseases, estimated at 8-20% [95]. Guidelines recommend active post-transplant surveillance for HCC patients, such as regular liver imaging tests within the first postoperative year and subsequent monitoring to detect lung metastases [96]. Tumor recurrence in HCC patients after transplantation is often attributed to advanced tumor burden and unclear tumor biology [97].

The Milan criteria, comprising specific size and number requirements for liver lesions along with the absence of vascular invasion or extra-hepatic metastases, were established to guide liver transplantations for Hepatocellular Carcinoma (HCC) [98]. Transplantations adhering to these criteria have demonstrated comparable survival outcomes to those performed for cirrhosis. However, criticism of the Milan criteria centers on their strictness in terms of lesion size and number, with some studies suggesting successful transplantation outcomes for HCC patients beyond these criteria. Additionally, the Milan criteria do not account for tumor biology, potentially limiting their applicability [99].

Down-staging, a strategy involving loco-regional therapy (LRT) to reduce tumor burden and bring lesions outside the transplant criteria within the criteria, has shown promise in achieving favorable long-term outcomes for HCC patients beyond the Milan criteria. Nevertheless, tumor recurrence remains a concern, occurring in 8-20% of transplanted HCC patients, typically within 2 years post-transplantation, with a median survival of 1 year following recurrence diagnosis [100].

To address the risk of tumor recurrence, various prognostic scores have been developed, such as the Risk Estimation of Tumor REcurrence After Transplant (RETREAT) score. This score considers three factors associated with post-transplant HCC recurrence: explant liver tumor burden, microvascular invasion evidence, and AFP levels at the time of transplant. The RETREAT score ranges from 0 to 8, with higher

scores indicating an elevated risk of recurrence. A score of 0 corresponds to a 1% recurrence rate at 1 year and a 2.9% recurrence rate at 5 years. Conversely, RETREAT scores of 5 or higher are associated with 1- and 5-year HCC recurrence rates of 39.3% and 75.2%, respectively [101]. Deep learning models can be used for diagnosis of HCC [102, 103].

The RETREAT score, while valuable for post-transplant management, has limitations as it relies on factors that assess explant tissue biology and anatomy. This restricts its utility to assessing transplant failure risk after transplantation. Machine learning (ML) models, utilizing pre-transplant data in HCC patients, can effectively allocate liver allografts before transplantation, thereby enhancing long-term survival prospects [101].

Although ML is gaining traction in various medical disciplines, this review reveals a dearth of pediatric studies among the 23 studies discussing ML and liver transplantation. This shortage reflects the limited interest in applying ML in pediatric patients, aligning with trends in other pediatric disciplines where ML adoption has been low. Consequently, there's a clear need for more research on ML in pediatric liver transplantation to assess its impact in this domain [104]. Furthermore, the high mortality rate in pediatric acute liver failure underscores the importance of robust criteria, including ML models, to inform decision-making in this patient group [105].

Evaluating ML model performance involves various metrics like accuracy, precision, confusion matrix, recall, specificity, precision-recall (PR) curve, F1 score, and receiver operating characteristic (ROC) curve. The use of ROC values in this study for assessing different ML models across studies is justified and reliable.

The utilization of ML algorithms in LT prognostication is a significant advancement. These models are primarily based on pre-transplant donor and recipient data, allowing for accurate predictions before transplantation. Considering that crucial decisions regarding liver transplantation must be made pre-procedure, ML models hold promise in addressing the complex challenge of allocating allografts to the most suitable recipients [101].

Numerous studies reviewed consistently indicate that ML models provide satisfactory to excellent predictions for both short- and long-term mortality or complication risks [106]. Additionally, emerging evidence suggests that artificial intelligence can surpass traditional tools in predicting cardiac events post liver transplantation [107] and mortality related to esophageal variceal bleeding [108, 109]. Accurate predictions of short- and long-term complications following liver transplantation are crucial, as they inform the need for additional surveillance or even potential halting of the transplantation process for patients at higher risk of mortality. Long-term complications post liver transplantation remain a significant concern, with limited improvement in survival rates over the years [110].

Long-term survivors face increased risks of comorbidities like metabolic syndrome, renal dysfunction, cardiovascular disease, and extrahepatic malignancies, necessitating multidisciplinary management strategies to prevent medical complications and their associated cost implications [111]. Metabolic syndrome, in particular, is prevalent among liver transplant recipients and is associated with chronic liver disease progression and increased cardiovascular risk [110]. Sustained transient post-transplant diabetes significantly elevates the long-term risk of major adverse cardiac events and mortality [113]. Therefore, precise prognostication of patients at risk of long-term complications is essential, and AI algorithms offer promise in enhancing risk assessment and improving patient outcomes.

Furthermore, ML models consistently outperform traditional scoring systems, including MELD, D-MELD, SOFT, p-SOFT, BAR, DRI score, ABIC, CLIF-C OFs, CLIF-C ACLFs, and CLIF SOFA, as well as models based on Cox and logistic regression. This finding is particularly significant given the limitations of traditional scoring systems in predicting post-transplant outcomes [101]. The incorporation of ML algorithms in organ allocation can enhance efficiency by preventing unnecessary transplantations and allocating allografts to patients with a higher likelihood of success. This optimization helps manage the associated costs of transplant failure and complications, especially considering the limited availability of donor organs. Regarding short and long-term

mortality prediction (90-day), the Random Forest (RF) model consistently exhibits the highest area under the curve [114] [115].

ML models provide numerous advantages, such as managing large datasets, objectivity, and assisting in cases with similar probabilities. In liver transplantation, artificial neural networks (ANNs) and random forest classifiers are the commonly used artificial intelligence models. ANNs excel at identifying complex patterns beyond human capability and can yield near-perfect predictions, reaching up to 95% accuracy in 3-month graft survival. However, ANNs lack transparency regarding the variables they consider. In contrast, random forest (RF) models offer better confidence in utilizing marginal organs, resulting in improved post-transplantation outcomes [114].

RF models exhibit superiority when predicting the risk of sepsis and acute kidney injury (AKI). Although overall survival post-liver transplantation has improved, post-transplantation infections remain a significant challenge, contributing to morbidity and mortality. Studies reveal that 35-55% of liver transplant recipients experience infection-related complications, including bacterial, fungal, and multidrug-resistant infections. Most of these infections occur within the first six months after transplantation and are responsible for a significant portion of early post-transplant deaths [116-119].

AKI and chronic renal dysfunction are common complications following liver transplantation. Contributing factors include long-term exposure to immunosuppressive medications like calcineurin inhibitors (CNI), preoperative kidney dysfunction, perioperative acute kidney injury/hypertension, diabetes mellitus (DM), and atherosclerosis pre- and/or post-transplantation. Long-term data indicates that kidney failure, defined as a glomerular filtration rate of 29 mL/minute/1.73 m^2 or less or the development of end-stage renal disease (ESRD), occurs in 18% at 5 years and 25% at 10 years post-transplantation [120]. Factors significantly associated with worse survival in patients with renal dysfunction include higher age at transplantation, increased creatinine levels, post-transplant diabetes mellitus, and transplantation in the pre-MELD era. Consequently, serum creatinine was incorporated into the MELD score to prioritize donor livers for transplant candidates with renal dysfunction [121] [122]. AKI

immediately following liver transplantation is linked to increased morbidity and mortality, with an incidence ranging from 25% to 60% [95].

The use of ML models in predicting the risk of sepsis and AKI is vital to enhance post-liver transplant outcomes. Post-transplant infections and AKI are associated with increased healthcare costs, prolonged hospital stays, and adverse effects on both allograft and patient survival [116] [119]. Also, ML models have been used for the diagnosis of appendicitis and heart disease [123] [124]. Employing ML models for predicting and managing these complications holds the potential to yield improved patient outcomes, reduced healthcare expenditures, and an overall better quality of life.

Despite the demonstrated superiority of ML models in the review, certain limitations must be acknowledged. Many studies relied on retrospective designs, which can introduce biases and impact result generalizability. Prospective studies with larger sample sizes and more diverse populations are necessary to validate ML model performance across different contexts and patient groups.

Another limitation stems from the lack of standardization and consistency in data collection and reporting of liver transplantation-related variables across various centers and studies. Data collection disparities can result in inconsistencies and hinder accurate comparisons of different ML models. Efforts should be made to standardize data collection practices in liver transplantation research to enhance the reliability and general applicability of ML models.

The underrepresentation of pediatric liver transplantation in the reviewed studies underscores a research gap. Pediatric patients have unique considerations and challenges in liver transplantation, and developing ML models tailored to this population could significantly enhance their outcomes.

Ethical considerations are paramount when implementing ML models in clinical decision-making. These models must be transparent, explainable, and accountable to ensure that clinicians and patients comprehend the rationale behind predictions, enabling informed decisions. Furthermore, addressing the black box dilemma of AI

models for prognostication is imperative, as ensuring transparency and interpretability in these models is essential to uphold ethical standards in healthcare decision-making.

CONCLUSION

CONCLUSION

This study reveals a significant surge in interest in the application of machine learning (ML) for liver transplant prognostication, with the majority of the studies emerging within the past five years. Notably, the United States and China stand out as the frontrunners in this field. This research also emphasizes that the performance of ML models exhibits variability when applied across different countries, underscoring limited external validity. Consequently, ML algorithms tailored to each country's unique transplant registry data demonstrate greater reliability.

Furthermore, the study highlights the superior predictive accuracy of ML models built on pre-transplant data in comparison to established scoring systems like MELD, irrespective of the underlying cause of hepatic failure, including HCC. Additionally, the study suggests that when selecting an ML model for predicting the risk of sepsis and acute kidney injury post-liver transplantation, the RF model may be the most suitable choice.

Overall, the use of ML models in liver transplantation has the potential to optimize organ allocation, improve patient outcomes, and reduce healthcare costs. However, more prospective studies with larger and diverse populations are needed to validate ML model performance and standardize data collection practices in liver transplantation research. Additionally, the inclusion of pediatric patients in ML research is crucial to address their unique needs. With continued research and advancements in ML techniques, ML models are poised to play an increasingly pivotal role in liver transplantation in the coming years.

ARTICLE HIGHLIGHTS

Research background

+ADw-html+AD4APA-p+AD4-Liver transplantation (LT) is a life-saving procedure for individuals with end-stage liver disease, offering not only health restoration but also a potential 15-year extension of life. However, the equitable allocation of donor organs remains a challenge due to donor scarcity. While the survival rates post-transplant are commendable, the shortage of donor organs persists, pushing the field towards utilizing less conventional donors. An efficient system of liver organ allocation is essential as there's a growing demand, leading to escalating healthcare costs. Traditional scoring systems like Child-Pugh-Turcotte (CTP) and Model for End-stage Liver Disease (MELD) have been employed for organ allocation, but they have limitations, such as empirical variable selection and limited predictive ability.+ADw-/p+AD4APA-/html+AD4-

Research motivation

+ADw-html+AD4APA-p+AD4-The primary challenge in liver transplantation is optimizing organ allocation. The scarcity of donor organs necessitates accurate prognostication for organ allocation and transplant success. While traditional scoring systems have been useful, they are not without limitations. Therefore, there's a need to explore more reliable and predictive methods. In this context, machine learning (ML) models present a promising avenue. ML algorithms can analyze various data types, from structured to unstructured, and offer a new dimension in predictive accuracy. Their ability to handle complex datasets and discover intricate patterns makes them suitable for enhancing prognostication in liver transplantation. Given the critical importance of optimizing organ allocation and predicting transplant outcomes, evaluating the utility of ML models is a significant step towards improving the liver transplantation process.+ADw-/p+AD4APA-/html+AD4-

Research objectives

+ADw-html+AD4APA-p+AD4-The primary objectives of this study are to comprehensively assess the effectiveness of ML models in liver transplantation

prognostication and to compare their performance and reliability with traditional scoring systems. This evaluation involves a systematic review of observational studies to determine the real-world utility of ML models in predicting transplant outcomes. Realizing these objectives is crucial for advancing the field of liver transplantation and ensuring that patients receive the most suitable organs, ultimately improving survival rates and healthcare resource allocation. Moreover, the study aims to bridge the gap between machine learning and traditional scoring systems, shedding light on the potential of ML models to revolutionize prognostication in liver transplantation.+ADw-/p+AD4APA-/html+AD4-

Research methods

+ADw-html+AD4APA-p+AD4-This systematic review followed PRISMA-P guidelines and conducted a comprehensive literature search on PubMed/MEDLINE using specific terms related to machine learning, artificial intelligence, liver transplantation, and prognosis. It included all relevant observational studies without restrictions on publication year, age, or gender, focusing on machine learning models for liver transplantation prognosis and post-transplant complications. Exclusion criteria covered non-English papers, review articles, case reports, conference papers, studies with missing data, or methodological flaws. A single reviewer screened and analyzed eligible studies, summarizing their objectives, methods, results, and conclusions. Data extraction included study type, population, year, purpose, setting, methods, results, and strengths/Limitations. The review also compared machine learning models to traditional scoring systems. This systematic approach synthesized information, offering a comprehensive understanding of artificial intelligence's role in liver transplantation prognosis and identified trends and potential benefits and limitations. It provides valuable insights into the current state of research in predicting liver transplantation outcomes with AI.+ADw-/p+AD4APA-/html+AD4-

Research results

In this systematic review, an initial pool of 64 references was identified and refined through a selection process. After excluding conference articles, review papers, and duplicates, 23 studies were included for analysis. These studies spanned from 2014 to 2023 and covered various transplantation reasons, with the majority conducted in the United States (34.8%), followed by China (26%). The primary outcomes assessed were mortality and post-transplant complications, with machine learning models consistently outperforming traditional models and scoring systems. The receiver operating characteristic (ROC) curve analysis demonstrated machine learning models' excellent predictive performance for both short-term and long-term outcomes. Notably, the Random Forest (RF) model excelled in predicting 90-day mortality, while the 'gradient boosting machine' model showed proficiency in forecasting complications like graft-versus-host disease, pneumonia, and acute kidney injury. The RF model was particularly adept at predicting sepsis and AKI. These findings highlight the potential of machine learning to enhance prognostic accuracy and inform clinical management in liver transplantation.

Research conclusions

+ADw-html+AD4APA-p+AD4-This study underscores the growing interest in applying machine learning (ML) to liver transplant prognostication, with a surge in research within the last five years. Notably, the United States and China have been leaders in this field. The research emphasizes the need for customized ML algorithms, adapted to each country's unique transplant registry data, to enhance the reliability of predictions. ML models, based on pre-transplant data, consistently outperform established scoring systems like MELD, regardless of the underlying cause of hepatic failure, including HCC. Additionally, when selecting an ML model for predicting the risk of sepsis and acute kidney injury post-liver transplantation, the Random Forest (RF) model appears to be a promising choice. These findings point to the potential of ML models in optimizing organ allocation, improving patient outcomes, and reducing healthcare costs in liver transplantation.+ADw-/p+AD4APA-/html+AD4-

Research perspectives

The future of research in this field should focus on conducting more prospective studies with larger and diverse patient populations to validate the performance of ML models and enhance their generalizability. Standardizing data collection practices in liver transplantation research is crucial to ensure consistency and facilitate accurate comparisons of different ML models.

Furthermore, there is a pressing need to include pediatric patients in ML research to address their unique requirements and challenges in liver transplantation. Ethical considerations should remain paramount, with a focus on ensuring transparency, explainability, and accountability in ML models to uphold ethical standards in healthcare decision-making. Continued advancements in ML techniques and the expansion of research efforts are expected to play an increasingly pivotal role in liver transplantation, offering the potential to further enhance patient care and clinical decision-making in the coming years.

ACKNOWLEDGEMENTS

We extend our appreciation to the Faculty of Life Sciences and Education at the University of South Wales for the Gastroenterology MSc program and their invaluable support in our work. We sincerely acknowledge the efforts of the University of South Wales and commend them for their commitment to providing life-long learning opportunities and advanced life skills to Healthcare professionals.

88891_Auto_Edited.docx

ORIGINALITY REPORT

5%
SIMILARITY INDEX

PRIMARY SOURCES

1	www.ncbi.nlm.nih.gov Internet	96 words — 1%
2	www.spandidos-publications.com Internet	40 words — < 1%
3	www.xiahepublishing.com Internet	32 words — < 1%
4	Eva María Pueyo-Périz, Guiomar Fernández-Castellano, Sonia Arnal García, Juan Luis López Romero et al. "Impact of the Acute and Chronic Liver Failure–Sequential Evaluation of Organ Failure Scoring System on the Prognosis of Liver Transplant", Transplantation Proceedings, 2022 Crossref	22 words — < 1%
5	Georges-Philippe Pageaux. "Mycophenolate mofetil in combination with reduction of calcineurin inhibitors for chronic renal dysfunction after liver transplantation", Liver Transplantation, 12/2006 Crossref	22 words — < 1%
6	Elahe Ahmadnia, Tayebe Ziaei, Masud Yunesian, Arezoo Haseli, Ali Mohammad Nazari, Afsaneh Keramat. "Sexual compatibility and its associated factors among heterosexual couples: a systemic review", Sexual and Relationship Therapy, 2021	20 words — < 1%

Crossref

7	www.researchgate.net Internet	17 words — < 1%
8	www.researchsquare.com Internet	17 words — < 1%
9	export.arxiv.org Internet	16 words — < 1%
10	link.springer.com Internet	16 words — < 1%
11	hepatmon.com Internet	15 words — < 1%
12	www.mdpi.com Internet	15 words — < 1%
13	www.frontiersin.org Internet	14 words — < 1%
14	www.science.gov Internet	14 words — < 1%
15	pjms.org.pk Internet	13 words — < 1%
16	d-scholarship.pitt.edu Internet	12 words — < 1%
17	speedypaper.x10.mx Internet	12 words — < 1%

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