Artificial Intelligence and Liver Transplant: Predicting Survival of Individual Grafts
A systematic review

Word count: TBD

Dr. Laura Wingfield[1], Dr. Carlo Ceresa[1], Simon Thorogood[2], Dr. Jacques Fleuriot[2], Dr. Simon Knight[1]

1. Nuffield Department of Surgical Sciences, University of Oxford, Room 6607, Level 6, John Radcliffe Hospital, Headington, Oxford, OX3 9DU, UK
2. The School of Informatics, University of Edinburgh, Informatics Forum,10 Crichton Street,Edinburgh,EH8 9AB, UK

Key words: Machine learning, hepatic, organ allocation, artificial neural networks, patient outcomes

Abstract

Background:

The demand for liver transplantation far outstrips the supply of deceased donor organs, and so listing and allocation decisions aim to maximise utility. Most existing methods for predicting transplant outcomes utilise basic methods such as regression modelling – newer artificial intelligence techniques have the potential to improve predictive accuracy.

Aims: To systematically review studies predicting graft outcomes following deceased liver transplantation using Artificial Intelligence (AI) techniques and comparing these to linear regression and standard predictive modelling (donor risk index, DRI; Model for end-stage liver disease, MELD; survival outcome following liver transplantation, SOFT).

Methods: A systematic review was performed. PubMed, Cochrane, MEDLINE, Science Direct, Springer Link, Elsevier, and reference lists were analysed for appropriate inclusion.

Results: A total of 52 papers were reviewed for inclusion. Of these papers, 9 met the inclusion criteria, reporting outcomes from 18,771 liver transplants. Artificial neural networks (ANN) were the most commonly utilised methodology, being reported in 7 studies. Two studies directly compared Machine Learning (ML) techniques to liver scoring modalities (i.e. DRI, SOFT, BAR). Both of these studies showed better prediction of individual organ survival with the optimal ANN model reporting AUC ROC 0.82 compared with BAR: 0.62 and SOFT: 0.57; and the other ANN model showing an AUC ROC: 0.84 compared to DRI: 0.68 and SOFT: 0.64.

Conclusion: AI techniques can provide high accuracy in predicting graft survival based on donors and recipient variables. AI approaches need to be integrated into current organ allocation methodologies to ultimately create better individual patient outcomes and will personalise decisions to organ acceptance.

Introduction

The number of patients awaiting liver transplantation currently outnumbers the number of donor livers available, which results in up to 20% of patients dying whilst on the waiting list.[1,2] As this is a scarce resource, it is becoming increasingly important to increase liver graft utilization and, at the same time, ensuring that the best possible outcomes can be achieved. This outcome paradigm is heavily reliant on a number of complex factors between donor, recipient and healthcare components.[3,4] Many different allocation systems have been utilised to find an appropriate solution to the delicate balance of factors needed to predict the best outcomes in liver organ allocation.

The current gold standard in prioritising patients waiting for a liver in throughout much of the world, remains the model for end-stage liver disease (MELD). However, this scoring system can have conflicting results.[5,6] Other liver scoring systems include the balance of risk score (BAR) and the survival outcome following liver transplantation (SOFT), which have been validated and are currently used to assist a surgeon in the decision-making process.[7] Many current scoring systems focus on the recipient survival prognosis, which is believed to be an extremely complex relationship that is non-linear in nature.[8] However, the majority of current liver allocation models utilised globally use methodology such as multiple regression and other linear models.[9] Therefore, achieving the most reproducible model which takes into account all donor and recipient factors would undoubtedly improve both organ utilisation and outcomes.[10,11,12]

The key to optimal organ allocation in transplantation is accurate prediction of an individual transplant outcome for a given set of donor and recipient variables. Such a

prediction can be used as part of matching algorithms to maximise overall benefit from the available organ pool. It can also be used during the patient-clinician decision-making process when deciding whether to accept an organ offer.

As clinicians search for better and more accurate models predicting transplant outcomes, newer technology is being trialled in the matching of this scare resource. Artificial Intelligence (AI), an area of computer science that encompasses intelligent and learned behaviours in computing, has been applied to many fields to produce better and more meaningful data analysis (Figure 1).[13] Specifically, Machine Learning (ML), a branch of AI that extrapolates patterns and information from provided data without necessarily being explicitly instructed to do so, is quickly emerging as a vital tool in the surgical sciences for outcome prediction.[14,15]

Machine learning itself can be further categorised into different sub-groups that include knowledge-based, supervised, unsupervised and reinforcement learning methodologies.[16] For the task of predicting outcomes following a liver transplant, supervised machine learning approaches are used which employ a combination of previously observed covariates (in this case, donor and recipient factors) and outcomes (in this case, observed survival times) to learn underlying relationships.

The distinction between prediction using ML and traditional statistical inference is not entirely clear-cut, but for the purposes of this review we differentiate as follows (Figure 1). Statistical models are those that assume an underlying probability distribution for the

data generating process. A relatively small number of parameters that define the distribution are then estimated from samples of data. In the field of survival analysis for example, Cox proportional hazards regression is a commonly used statistical model.[17] In contrast, ML approaches are more algorithmic in nature and typically have much larger numbers of associated parameters. These approaches allows complex non-linear interactions between factors to be learnt directly from data samples with few, if any, assumptions being made about the underlying distribution of the data generating process itself. Some examples of ML techniques that are commonly used for predictive tasks are artificial neural networks (ANNs), random forests (RF) and support vector machines (SVMs).[16] ANN are models that use principles of statistics to build complex modelling tools using data that is nonlinear in nature, and it imitates human thinking in the way it process several data types and creates patterns that are ultimately used in decision-making through these neural networks.[18] RFs create multiple decision trees that are able to sort through data and identify important variables that influence prediction or outcome.[19] Finally, SVM methodology organises data by class of variables (in a non-linear modality), known as hyperplanes that are able to form complex multi-dimensional, infinite planes in space utilising this data.[19]

To our knowledge, this is the first systematic review of AI computing techniques being used in liver transplantation to predict individual patient graft survival. The aim of this research is to provide a review of the collective evidence on AI computing techniques to predict individual patient liver graft survival when compared to the standard risk scoring tools (MELD, DRI, SOFT).

Methods

Literature Search Strategy

Original studies on Artificial Intelligence used to predict individual patient liver graft survival were identified by searching the following databases: MEDLINE, Science Direct, Springer Link, Elsevier, PubMed Central, and Cochrane databases from inception to 11 September 2019. ClinicalTrials.gov was also searched for relevant ongoing trials. No date limitations were included within the search parameters. The following keywords were used in the search: MeSH terms including 'Machine Learning' OR 'Artificial Intelligence' OR 'Neural Networks (Artificial)' OR 'Support Vector Machine' OR 'Stochastic Processes' OR 'Bayesian Learning' OR 'Supervised Machine Learning' OR 'Machine learning[title/abstract] OR neural network [title/abstract] OR Bayesian Learning [title/abstract] OR support vector [title/abstract] OR machine learning [title/abstract] OR deep learning [title/abstract] OR 'stochastic processes'[title/abstract] AND MeSH terms 'Hepatic Transplantation' OR 'Liver Transplantation' OR 'Liver Grafting' OR 'Hepatic Transplant*' [title/abstract] OR 'Liver Transplant*'[title/abstract] OR 'Liver Graft*'[title/abstract] NOT editorial[publication type] or comment[publication type]. The literature search included studies published in any language. Additionally, a manual review of the reference lists of the studies obtained from the above described search strategy was used to identify additional relevant studies. PRISMA Checklist was utilised to include appropriate studies within this review. The study was registered on the PROSPERO database (CRD42019094865).

Selection Criteria

Studies were included if they looked at patients who received a deceased donor liver transplant with the intervention of AI computing techniques with traditional statistical modeling to determine individual graft survival outcomes. Studies were limited to those including only liver transplant, and those studies that involved AI and renal, cardiac, or lung transplants were excluded. Only adult transplant patients were included. Modalities of AI included within this review encompass mainly neural networks, random forests and probabilistic graphical modelling. Overlapping study groups, although containing the same data sets, were included if they utilised distinctly different types of ML techniques to analyze the data or looked at different survival timescales. If studies looked at other disease areas around transplantation (i.e. hepatic carcinoma recurrence post-transplantation), these studies were excluded.

No limits or exclusions were made on the number of transplant recipients or the country location of the transplants. Study design types within the inclusion criteria for this review were meta-analysis (of randomised control trials), randomised control trials and cohort studies. Data extraction to identify ML methods, ML methodology and associated results were recovered via a data extraction sheet. This data extraction was conducted by two independent researchers (LW and CC) and any discrepancies were resolved by the senior author SK.

Quality assessment Methods

The studies within this review were assessed for quality utilising the Critical Appraisal Skills Programme (CASP) Checklist for cohort studies.[20] A table was constructed (Table 1) which details the four broad areas of quality assessment include 'validity of results,' 'worth continuing work,' 'results,' and 'results helpful locally.' The table assessed each corresponding item in the above mentioned categories by either 'yes', 'no,' or 'unsure.'

Results

Literature review

Of the initial 52 studies identified via the search terms, 24 papers were excluded as they did not specifically include liver transplant, or they did not include AI. The remaining 28 papers were selected for a further full manuscript review. From this group of papers, eight papers were excluded as upon further examination they did not include pure AI modelling, and a further two papers that focused solely on testing the feasibility of computing models but did not provide outcome data were excluded. For example, the paper by Tusch focused on a patient decision making pathway pre and post-transplant.[21] A further 10 papers were excluded because although they examined AI modalities, they looked at it in association to transplant related disease (i.e. recurrent of hepatocellular carcinoma in post-transplant patients), not as a tool for prediction of individualised graft survival outcomes. One paper examined paediatric liver transplant survival, therefore, it was excluded. A final 9 papers [22,23,24,25,26,27,28,29,30] were selected for inclusion within this study. [Figure 2, PRISMA Flow Diagram Paper Selection]

The 9 papers relevant to the topic of AI and liver transplant outcomes, included a total of 18,771 study participants [Table 2]. Within these papers, the majority examined graft survival at 3 months and / or graft survival at 1 year with eight papers examining these end points each. One paper examined survival between 2 to 5 years. At the other end of the spectrum, some articles looked at graft survival in the very short term with 2 papers reviewing survival predictions at 30 days or less.

Participants

There was a mean of 2,086 (range: 180 to 12,239) participants from all studies included in the review and there was a tendency to analyse larger cohorts of data. For example, Hoot *et al.* reviewed a cohort of 12,239 patients from American UNOS registry data.[27] Additionally, groups were analysed from a number of countries and transplant data sources including Spain, UNOS data from the United States, several UK centres, an Australian Liver Database, and China. [Table 2] Notably, five of the study groups utilized the same dataset in their research which was comprised of Spanish transplantation centers with an externally validated dataset from King's College, London.[22,23,24,25,29]

Quality Assessment of studies included in the review

The papers included were good quality overall [Table 1], with all papers reviewed including a clear research aim. A total of 77% of the papers reviewed contained results that could be used in clinical practice (n=7). Despite these strong quality points, there were other areas that the papers could have improved upon. For instance, in almost all of the papers reviewed (n=8), it was difficult to determine whether the results could be

applied to local populations. Further areas of improvement were centered around bias reporting. Although the authors within the studies described may have specifically measured outcomes to minimise bias, this was not explicitly mentioned in three papers.[26,27,29]

Artificial Intelligence Approaches and Input Features

There is currently no gold standard in AI modelling for clinical outcomes and a number of different approaches were trialed within our study groups. Approaches utilised within the articles reviewed for this paper include: artificial neural networks (ANNs), support vector machines (SVM), random forests, gradient-boosted trees and Bayesian networks. More than half of the studies included (n=5) trialed more than one type of ML approach to predict the same outcome (i.e. graft survival).[23,24,25,28,29] Despite the survival analysis setting and the existence of right-censored data, most studies approached the task as a binary classification problem at a specific time (e.g. graft survival at 1-year post-transplant). As such, observations censored prior to the time of interest are excluded from the modelling process. A number of articles approached the task as one of ordinal classification (i.e. the simultaneous predictions of 2 or more intrinsically ordered classes) where the classes represent failure at successive times of interest. The most commonly used type of classifier was ANNs, with 7 papers adopting this approach [Table 3].[22,23,24,25,28,29,30]

It is well established that the process of organ allocation is extremely challenging and can be influenced by many potentially interacting factors. Within the 9 papers included, there was a range of donor and recipient characteristics that were examined. The

number of variables ranged from 10 to 276 donor, recipient and surgery-related ones. The Lau group measured the largest number of overall variables: 173 recipient variables and 103 donor recipient variables initially. They reviewed these variables and selected a 'top' 15 inputs following feature selection.[28]

Comparisons between AI modelling and standard organ risk stratification systems were made within almost all the papers included. Standard liver risk stratification scores used for comparison include MELD, SOFT, P-SOFT, BAR and DRI, with MELD being the most commonly used (n=8). Where comparisons were quantitative, a number of performance metrics were used. The most commonly used was the area under the receiver operator curve (AUC ROC)[22,23,24,27,28] with C-index[30], accuracy, sensitivity and the geometric mean of sensitivities (GMS)[29] also reported in some cases. When reviewing the included manuscripts, it is important to contextualize the results in reference to the AUC, or c statistic, which range from 0.5 (showing no discrimination) to a perfect model showing a maximum value of 1. Therefore, an AUC score of 0.5 would be equivalent to a coin flip of chance. In a clinical context, models with an AUC score of 0.7 are considered a 'good fit' with an AUC score of 0.9 being an almost 'perfect' model.[31] Ultimately, clinical judgement was utilized in combination with standard organ risk stratification systems including MELD scoring and previous UNOS allocation systems.

Validation of AI models used

Within the papers selected for this review, there were a variety of approaches undertaken to validate the data, however all 9 studies validated their datasets. The most common methodology was cross validation (n=5) with researchers utilizing either 3-fold, 5-fold stratified cross-validation or 10-fold stratified cross-validation [Table4].[22,25,25,29,29] Another approach taken was a training and test data set with Zhang's group including a 80%/20% train-test split with additional 20% validation set created from training set, and Cruz-Ramírez including a 75%/25% train-test split, with multiple bootstrap samples created from each set.[23,30] Lau utilized a 1,000 bootstrap samples with out-of-bag samples for validation and was the only group to take this approach most likely due to their small sample size.[28] Interestingly Haydon *et al.*, did not explicitly mention within their work the parameters utilised in evaluating the efficacy of their model, only mentioning that a separate database of 2,622 patients was used for validation, and were the only group not to mention these.[26]

AI compared with other predictive methods

Five of the studies directly compared the performance of ML models to some form of linear regression modelling [Table 3].[21,23,25,27,28,29] Almost half of the studies within this review did not mention specific regression methodology used (n=4). [24,26,29,30] Of the studies that directly compared ML models with standard regression modelling (n=5) only two out of this total reported higher accuracy in ML. For instance, Brienco et al. reported their best ML model as Artificial Neural Networks (ANN) with accuracy: 0.91 and AUC ROC: 0.82 compared with their most successful logistic regression model with

accuracy 0.89.[22] Although statistically speaking this is an improvement, an improvement of 0.02 between traditional logistic and ML models cannot be considered a clinically significant improvement. Cruz-Ramirez also compared ML techniques to logistic regression and showed similar accuracy between the two modalities but overall improved AUC ROC with ML. The logistic model tree ML technique had an accuracy of 0.88 while the best logistic regression model was 0.88, the ANN model showed minimum sensitivity 0.50 compared to logistic regression of 0.03, and AUC ROC: 0.57 compared with logistic regression AUC ROC: 0.51.[24]

Finally, only two studies directly compared ML techniques to liver scoring modalities (i.e. DRI, SOFT, BAR). Brienco *et al*. showed an ML AUC ROC: 0.82 compared with BAR: 0.62 and SOFT: 0.57.[22] The 0.20 difference in AUC between traditional and ML models can considered clinically significant. Following from these results, the Brienco study group suggests that their ML scoring system utilizing ANN could be used to predict 3-month outcomes in conjunction with clinical judgement. They also state that they believe that ANNs may be the best method to combine the myriad of variables involved in transplantation (i.e. donor, recipient and others) to obtain optimal survival. They do not directly state that this methodology should replace current liver indices or linear regression models. However, they do explain that many of the current scoring modalities (DRI, MELD, SOFT, and BAR) utilize logistic regression analysis which assumes a 'linearity' amongst the liver transplant variables and survival. The authors point out that in reality, liver transplantation follows a nonlinear patter and therefore this approach is too simplistic. The other research to directly compare ML modalities to

traditional liver scoring were the Lau research group. The Lau ANN model showed an AUC ROC: 0.84 compared to DRI: 0.68 and SOFT: 0.64.[28] These results show a difference of 0.16 AUC values between the best ML model and the DRI, which can be considered high enough to practically warrant clinical use of one model over another. Despite Lau's results, the study group note that this initial research was a proof of concept research that could potentially be used to support clinical decision making in liver transplant organ allocation. They do not outwardly suggest that this methodology should be used instead of current liver indices or linear regression modelling. However, Lau suggests that their ML algorithm could be used as a tool to improve clinician confidence in using marginal organs.

Discussion

To our knowledge, this systematic review is the first of its kind that reviews ML methodology to predict individual graft outcome following liver transplantation. Other systematic reviews have examined ML and renal graft outcomes as reported by Senanayake *et al.*, Nursetyo *et al*., and Sousa *et al.* reviewed ML applications in heart, heart-lung, and kidney transplanted organs published between the years of 2009-2010.[32] These limited systematic reviews in AI/ML and organ transplantation highlight a gap in the literature.[33,34] As well as limited reviews in the topic, the results have highlighted the heterogeneity in the AI techniques used within the study results. Unfortunately, only a few papers directly compared ML modelling with logistic regression and / or liver scoring systems. It would be especially useful to examine this

parameter as researchers would be using the same data sets to compare ML performance with what is currently being used in liver scoring systems clinically.

Studies within this review by the Briceno, Cruz-Ramirez, and Lau groups demonstrated that ML modelling provided more accurate results when compared to standard regression and liver scoring modality. [22,24,28] Recently, in the United Kingdom, the liver allocation system has changed whereby a “transplant benefit score” is generated. This system aims to provide a more in-depth score than the previous UKELD classification and comprises 7 donor and 21 recipient characteristics and is calculated using linear regression modelling measuring the difference between the AUC for the waiting list survival curve and the AUC under the post transplantation five year survival.[35] In time, it will be important to establish whether this new approach results in an increase in the number of life-years gained from transplanted livers and decrease the number of waiting list deaths. Further studies are certainly needed to select a universal AI methodology and support from national bodies needs to be garnered. This review study highlights the growing evidence to support AI technology as a predictive tool that can be used to form an organ allocation system when compared to standard methods of allocation currently in use. Although AI technology is being used in other fields of medicine, it has yet to have widespread, nationally implemented programmes at the time of this publication.[36]

Advantages to ML Systems

In addition to the need for a nationally adopted standard AI system, the advantages of AI need to be further stressed with education of governmental and health organisations taking place. For instance, one advantage of AI methodology, especially neural networks, when compared to standard techniques is that they are dynamic – able to be trained and validated within every population. Furthermore, the more variables that are examined in terms of donor characteristics, the more precise a neural network can be, provided enough data is available. This allows better organ allocation decisions that take into account a large number of variables, which standard programmes currently may not include or may need a clinician's detailed review of donor and recipient parameters. In order to achieve this, however, there is a need for a standardized curation of liver transplantation data to allow widely usable, standardised models to be built to be used across different counties. Ultimately, for the individual patient, this information will help make more informative, personalized decisions around organ acceptance and adapt the informed consent process to the organ on offer at any one time. Finally, there are significant financial costs and regulatory constrains related to liver transplant and as these constrains increase, it is important to have a quantitative tool to help transplant clinicians make these critical organ allocation decisions.

Challenges with Implementing ML Systems

Despite potential benefits to utilizing ML in liver transplantation, there are also potential limitations to utilizing this emerging technology. Ultimately, researchers using ML-based algorithms aim to present the most accurate prediction output as the data will

allow. In some cases, the algorithms may include variables that based on clinical experience and previous research are not biologically plausible. This affords researchers, clinicians and the patients who will ultimately receive the liver transplant with a conundrum: if a variable is known to have no survival benefit in clinical practice, but it shows a survival benefit anyway, should it be included? This leads to the ethical issue around explainability in AI, which is described in detail below. Another limitation is that these algorithms may not have global applicability, instead they are often best suited to predicting outcomes based on datasets from which they were originally derived. Further challenges with ML algorithms may stem from shifting patient populations. As algorithms are designed based on relatively statistic datasets in a particular point in time, there may be a new distribution in the data compared to the original dataset used to train the ML algorithm. Finally, there are logistical challenges around translating ML algorithms directly into a clinical setting. For instance, different healthcare computer systems may not easily host the programs required to run the algorithms. Final hurdles to implementation may center around clinician acceptability and implementation of algorithms when there can be considerable opacity around the algorithms themselves.

Study Strengths and Limitations

The strengths of this review include that it is the first study to review the available literature of AI/ML techniques in liver transplant. Thus far there has been extremely limited work in this field, and this review aims to amalgamate the current work in this area to date. This review also discusses a timely subject, ML in healthcare, and as this

methodology is being rapidly integrated into healthcare systems it is vital that the research in this area is disseminated for the transplant community. Finally, this review covers a wide breadth and depth of study participants including transplant recipients from Australia, Spain, U.K., U.S.A. and China. A total of 18,771 liver transplantation patients comprised the datasets used by researchers also making it the largest, systematic review of its kind.

There are some limitations to this study. Many of the papers included within this review were observational studies – data being retrieved from databases. They were retrospective in nature, and in order to truly test the predictive power of AI, it would be ideal to have further large, externally validated prospective cohort studies. These studies could review prospective outcomes such as 30-day graft survival as well as longer-term graft survival rates (i.e. 1 year or more). An additional limitation to this study is the high rate of heterogeneity in AI techniques utilized amongst the papers included within this review as previously mentioned. The size, heterogeneity, and quality of datasets amongst the studies included make direct comparison between studies challenging including evaluations between regions / countries, early verses late prediction abilities of the ML algorithms and prediction comparisons by MELD strata groups. Finally, none of the studies took a time-to-event (i.e. graft failure) approach. In clinical practice it may be useful to have such a quantitative measure of the likelihood survival gain by accepting / rejecting a specific organ, which would help further clinicians and patients make more informed decisions on grafts.

Explainability in Machine Learning

ML methodologies are being used to analyse data in completely new ways. This evolution in science is creating a great potential to develop clinical decision support tools that can help doctors and patients make critical decisions about health. However, the use of ML in critical healthcare decisions brings up several challenges. The high accuracy of ML may come at a cost of losing explainability (to patients and clinicians) on how the technology works. Some ML-based algorithms work in ways that are unknown to the creator and therefore, cannot be explained to patients or doctors using them. (known as the 'Black Box' issue).[18] This raises questions about accountability for such algorithms in the event that an incorrect result (i.e. incorrect liver graft survival prediction) is made. Furthermore, incomplete datasets used to train ML algorithms may cause potential biased outcomes. Finally, the use of ML poses questions around the acceptability of patients and their careers when decisions are delegated (partially) away from humans and more so to computer algorithms. Based on these challenges around ML algorithms, it is essential to have research into how both clinicians and patients would interpret ML generated algorithms.

In spite of challenges around the "Black Box" issue, if AI techniques were widely accepted within liver transplant, this could be an invaluable tool in providing clinicians with critical decision making. It would allow surgeons to make more evidence-based decision-making as well as provide patients with a tool to understand the risk/benefit ratio of accepting a specific liver transplant. Finally, an AI model could make the liver donation system more efficient and would cause fewer organs to be discarded.

Acknowledgment

No acknowledgements were made within this manuscript.

Footnotes

Abbreviations

- **AI**

  Artificial Intelligence
- **ANN**

  Artificial Neural Network
- **AUC**

  Area under the curve
- **AUC ROC**

  Area under the receiver operator curve
- **ANN**

  Artificial Neural Network
- **DRI**

  Donor Risk Index
- **GA**

  Genetic Algorithm
- **MELD**

  Model for end-stage liver disease
- **ML**

Machine Learning

- **MLP**

  Multilayer Perceptron

- **PRISMA**

  Preferred Reporting Items for Systematic Reviews and Meta-analysis

- **RBF**

  Radial Basis Function

- **ROC**

  Receiver Operating Characteristic

- **SOFT**

  Survival outcome following liver transplantation

- **SOM**

  Self-organizing Map

Grants and Financial Support: The authors declare no grants or financial support in this work.

Conflict of interest: The authors of this manuscript have no conflicts of interest to declare

Mailing address for corresponding author:

Dr. Laura Wingfield, M.B.B.S.
Nuffield Department of Surgical Sciences,
University of Oxford, Room 6607, Level 6,
John Radcliffe Hospital, Headington, Oxford,
OX3 9DU,
UK
Email: laura.wingfield@ouh.nhs.uk
Tel: +44 776 06627 53

Appendix

Figure 1: Artificial Intelligence Defined
(see attached)

Figure 2: PRISMA flow diagram of systematic identification, screening, eligibility, and inclusion criteria.
(see attached)

Table 1: Critical Appraisal Skills Programme (CASP) Checklist for cohort studies

(see attached)

Table 2: Study demographics
(see attached)

Table 3: Results of Machine Learning methodology compared to regression and liver risk index scoring
(See attached)

Table 4: Study input variable and validation methodology
(See attached)

References

1. Neuberger J. Liver transplantation in the United Kingdom. Liver Transpl 2016;22:1129-1135.

2. Kim WR, Lake JR, Smith JM, Skeans MA, Schladt DP, Edwards EB, et al. OPTN/SRTR 2013 annual data report: liver. Am J Transplant 2015;15(suppl 2):1-28.

3. Busuttil RW, Tanaka K. The utility of marginal donors in liver transplantation. Liver Transpl. 2003;9:651–663.

4. Tector AJ, Mangus RS, Chestovich P, et al. Use of extended criteria livers decreases wait time for liver transplantation without adversely impacting posttransplant survival. Ann Surg. 2006;244:439–450.

5. Halldorson JB, Bakthavatsalam R, Fix O, et al. D-MELD, a simple predictor of post liver transplant mortality for optimization of donor/recipient matching. Am J Transplant. 2009;9:318–326.

6. Croome K, Marotta P,WallW, et al. Should a lower quality organ go to the least sick patient? Model for End-Stage Liver Disease score and Donor Risk Index as predictors of early allograft dysfunction. Transplant Proc. 2012;44:1303–6.

7. Flores A, Asrani SK. The donor risk index: A decade of experience. Liver Transpl. 2017 Sep;23(9):1216-1225.

8. Yeh H, Smoot E, Schoenfeld DA, Markmann JF. Geographic inequity in access to livers for transplantation. Transplantation. 2011 Feb 27;91(4):479-86.

9. Lewsey JD, Dawwas M, Copley LP, Gimson A, Van der Meulen JH. Developing a prognostic model for 90 day mortality after liver transplantation based on pretransplant recipient factors. Transplantation. 2006; 82: 898–907.

10 . Collett D., Friend PJ, Watson CJ. Factors Associated with Short- and Long-term Liver Graft Survival in the United Kingdom: Development of a UK Donor Liver Index. Transplantation. 2017 Apr;101(4):786-792.

11 . Braat AE, Blok JJ, Putter H, Adam R, et al. The Eurotransplant donor risk index in liver transplantation: ET-DRI. Am J Transplant. 2012 Oct;12(10):2789-96.

12. Feng S, Goodrich NP, Bragg-Gresham JL, Dykstra DM, et al. Characteristics associated with liver graft failure: the concept of a donor risk index. Am J Transplant. 2006 Apr;6(4):783-90.

13. Ghahramani Z. Probabilistic machine learning and artificial intelligence. Nature. 2015;521(7553):452-459.

14. Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. Science. 2015;349(6245):255-260.

15. Deo RC. Machine learning in medicine. Circulation. 2015;132(20):1920-1930.

16. Murphy K. Machine Learning: A Probabilistic Perspective. The MIT Press. 2012.

17. Cox, D. R. Regression Models and Life Tables (with Discussion). Journal of the Royal Statistical Society. 1972; 34:187—220.

18. Mintz Y, Brodie R. Introduction to artificial intelligence in medicine. Minim Invasive Ther Allied Technol. 2019 Apr;28(2):73-81.

19. Mortazavi BJ, Downing NS, Bucholz EM, Dharmarajan K, et al. Analysis of Machine Learning Techniques for Heart Failure Readmissions. Circ Cardiovasc Qual Outcomes. 2016 Nov;9(6):629-640.

20. Critical Appraisal Skills Programme (2018). CASP Cohort Checklist. [online] Available at: https://casp-uk.net/wp-content/uploads/2018/01/CASP-Cohort-Study-Checklist_2018.pdf Accessed: 1 October 2019.

21. Tusch G. An optimization model for sequential decision-making applied to risk prediction after liver resection and transplantation. Proc AMIA Symp. 1999:425-9.

22. Briceño J, Cruz-Ramírez M, Prieto M, Navasa M, *et al.* Use of artificial intelligence as an innovative donor-recipient matching model for liver transplantation: results from a multicenter Spanish study. J Hepatol. 2014 Nov;61(5):1020-8.

23. Cruz-Ramírez M, Hervás-Martínez C, Fernández C, Briceño J, *et al*. Multi-objective evolutionary algorithm for donor–recipient decision system in liver transplants. European Journal of Operational Research 222 (2012) 317–327.

24. Cruz-Ramírez M, Hervás-Martínez C, Fernández JC, Briceño J, *et al.* Predicting patient survival after liver transplantation using evolutionary multi-objective artificial neural networks. Artif Intell Med. 2013 May;58(1):37-49.

25. Dorado-Moreno M, Pérez-Ortiz M, Gutiérrez PA, Ciria R, *et al.* Dynamically weighted evolutionary ordinal neural network for solving an imbalanced liver transplantation problem. Artif Intell Med. 2017 Mar;77:1-11.

26. Haydon GH, Hiltunen Y, Lucey MR, Collett D, *et al.* Self-organizing maps can determine outcome and match recipients and donors at orthotopic liver transplantation. Transplantation. 2005 Jan 27;79(2):213-8.

27. Hoot N, Aronsky D. Using Bayesian networks to predict survival of liver transplant patients. AMIA Annu Symp Proc. 2005:345-9.

28. Lau L, Kankanige Y, Rubinstein B, Jones R. Machine-Learning Algorithms Predict Graft Failure After Liver Transplantation. Transplantation. 2017 Apr;101(4):e125-e132.

29. Pérez-Ortiz M, Gutiérrez PA, Ayllón-Terán MD, Heaton N, et al. Synthetic semi-supervised learning in imbalanced domains: Constructing a model for donor-recipient matching in liver transplantation. Knowledge-Based Systems. 2017; 123 (2017) 75–87

30. Zhang M, Yin F, Chen B, Li B, Li YP, Yan LN, Wen TF. Mortality risk after liver transplantation in hepatocellular carcinoma recipients: a nonlinear predictive model. Surgery. 2012 Jun;151(6):889-97.

31. Cook NR. Use and Misuse of the Receiver Operating Characteristic Curve in Risk Prediction. Circulation. 2007;115:928-935.

32. Sousa FS, Hummel AD, Maciel RF, Cohrs FM, et al. Application of the intelligent techniques in transplantation databases: a review of articles published in 2009 and 2010. Transplant Proc. 2011 May;43(4):1340-2.

33. Senanayake S, White N, Graves N, Healy H, et al. Machine learning in predicting graft failure following kidney transplantation: A systematic review of published predictive models. Int J Med Inform. 2019 Oct;130:103957.

34. Nursetyo AA, Syed-Abdul S, Uddin M, Li YJ. Graft Rejection Prediction Following Kidney Transplantation Using Machine Learning Techniques: A Systematic Review and Meta-Analysis. Stud Health Technol Inform. 2019 Aug 21;264:10-14.

35. NHS Blood and Transplant Organ Donation and Transplantation Directorate Advisory Group Chairs Committee (2014). Liver National Allocation Scheme. Fixed Term Working Unit – Organ Allocation.

36. Baxt WG. Application of artificial neural network to clinical medicine. Lancet 1995;346:1135–1138.