A Review: Machine Learning and Artificial Intelligence for Organ Transplant Survival Prediction

October 3, 2022

Abstract

Patient survival is the period between the time of organ transplant and the time of death. An increase of 0.1 in the index would have a significant impact on the organ allocation system. Consequently, it improves patient outcomes. There were several methods to predict survival; these techniques include conventional methods (linear regression) and modern methods (machine learning and artificial intelligence). However, the majority of patients and clinical practitioners in the system doubted the accuracy and interpretability of the latter. This review paper describes the significants of modern methods over traditional, in the context of survival. The motivation is to research the literature, present a statistical and informational report to help alleviate the distrust in machine learning and artificial intelligence for organ transplant survival prediction.

1 Introduction

More than 850 million in the world suffer from some form of organ disease. The demand for organ transplants (kidney, liver, heart-lung) is more than its availability. About 21 people die daily while waiting for a Renal Replacement Therapy (RRT) of their failing organ. Meanwhile, most organ transplant treatment does not last for a lifetime. The graft failure or graft survival time (the period until the replaced organ failed or needs re-transplant) varies based on widely varied risk factors that influence survival. Hence, there is a need to maximize the utility to have more graft survival time. There have been several linear methods used to predict the outcome of organ allocation. A percent increase in the precision and accuracy of outcomes predicted by a model is a big deal in the field of organ transplants. The more accurate the prediction, the less the number of re-transplant needed due to graft failure. The complex relationship between the donors, recipients, and health care providers influences the outcome. The recent electronic means of storing medical records have resulted in medical Big Data of various medical processes. Therefore, there is a need to use advanced techniques such as machine learning methods to make analyses since there are high computing machines. The Machine learning algorithms can improve their accuracy on predictions based on previous experience on data; would adapt to new data without redesigning the system, expose noisy data; are reproducible for use with as many factors as relevant. This work aims to present narrative information on the performance of machine learning methods to predict survival analysis in the field of organ transplants. This review paper is structured as follows:

The "Methods" section summarizes how each machine learning model was used in the selected documents.

The "Result" section presents the machine learning algorithms' performance measures compared to other methods.

The "Discussion and Conclusion" section discusses the findings, limitations, and future work ideas.

2 Literature Selection

This paper search for relevant literature using the google search engine and medical journals such as; BMC Medical Research Methodology, Medical Internet Research, Pubmed, and Cardiac Surgery. Some phrases and statements searched include human organ transplant, machine learning methods for survival analysis. This work has no restriction to the country location of data, age cohort, or specific organs. However, there is more attention to the size of data and the machine learning methodology used to ensure sufficient diversity in the study. The quality of this paper is based on the standard of their sources. We consider diverse methods in the choice of study participants, attributes used for modeling, modeling process, imputation method, individual learner, ensemble methods, and module structure. A summary of the selected papers are described in table 1 below:

Table 1: The choice of papers in this study.

Authors	**Data size**	**Organs**	**ML Methods**
Kantidakis et al.[1](2020)	62294	Liver	Partial Logistic ANN
Pahl et al.[2](2020)	70242	Kidney	Random Forest
Mark et al.[4](2019)	100000	Kidney	Decision Tree
Nemati et al.[5](2021)	100000	Kidney	Gradient Boosting
Naqvi et al.[7](2021)	50000	Kidney	Deep Learning
Don Yoo et al.[8](2017)	3117	Kidney	Survival Decision Tree
Ayers et al.[6](2021)	33657	Heart	Deep Neural Network
Killian et al.[9](2020)	1895	Kidney, Liver, Heart	Naive Bayes,

3 Methods

This section will describe how the algorithms listed in table 1 were used in their respective selected literature.

Partial Logistic ANN: An artificial neural network has been applied to survival analysis for years. For example, Kantidakis et al.[1] used prognostic factors and survival times as input to the partial logistic artificial neural network (PLANN), which increases the predictive accuracy of the model. The survival time was divided into a disjoint set, creating an interval of months and years. The transformation function for the input and output layer are functions of the discrete conditional probability of dying. Each disjoint survival time interval is treated as a variable, and the binary output vector describes the event in a time interval. The model's performance on the test data was assessed by measures the Integrated Barrier Score (IBS). The IBS measure takes a value between 0.5 and 1, and the neural network had a value of 0.180, which did not perform better than the other model used in their research work. It was reported to be a result of many weights used to train the model, which caused overfitting. More about the origin of the method can be seen in a work by Biganzoli E. et al.[3].

Random Forest: This algorithm is vastly used for the following reasons; it allows mixed data types with less or no data modification; it effectively handles outliers and co-linearity; it works well for the majority of healthcare data. Almost all the paper considered for this review uses Random Forest (RF) algorithm. Pahl et al.[2] use RF to create multiple decision trees to model the graft failure by randomly selecting a fraction of the training dataset. Each tree predicts maximizing the Gini index, and the RF outcome is a vote from each tree based on the majority vote. This approach is similar to getting an opinion from an expert based on experience and professional background. For this work, the result for IBS is 0.182, which is lower than some other models, but using the C-index measure, the impact of the RF model is 0.622, which performs better than the different models used in their work.
Mark et al.[4] used RF in the process of selecting which variables are essential for the model. RF raked age as of most important for their model, and the data was split based on age cohorts. The result of the RF model is better at (0.718) using the C-index measure while it does not use the IBS measure (0.062).

Gradent Boosting (GB): This method is used to boost weak learning algorithms by using survival regression trees. Nemati et al.[5] used a stochastic gradient with 200 trees, then conducted a grid search on the maximum depth of each tree in a five range interval. This technique shows a prediction accuracy of 0.639 for the C-index measures.

Deep Learning (DL): Naqvi et al.[7] in their work, applied deep learning with 13 -layers autoencoders to reduce the dimensionality of the data. Their method helped to increase the accuracy of the machine learning classification method used. Autoencoders would not take in categorical variables, so multiple dummies were created to convert 86 categorical variables. Then the stacked autoencoder with 12 dense layers and one drop-out layer was applied to the converted features and the continuous variables. More than 50 dummy features reduced the feature space after using 500 epochs and 700-900 batch sizes. A similar approach was employed with the work done on a heart transplant. Ayers et al.[6] reported that deep learning algorithm may have helped to defeat the registry-level data limitations that past studies had.

Survival Decision Tree: Yoo et al.[8] applied a survival tree algorithm with survival analysis statistics instead of the traditional Gini index. This method was used to identify an acute episode of rejection within the first year after as the most critical factor to predict graft failure. The survival tree algorithm performed better at 0.80 compared with the conventional method at 0.71.

Support Vector Machine: Most classification models pursued binary output like graft failure or survival. Naqvi et al.[7] used a support vector machine for classification by using kernels like radial, linear, sigmoid, and polynomial kernels. A polynomial of degree 2 was observed to perform better.

4 Results

In place of seeing numbers, it would have been nice to visualize how machine learning models performed compared with other models with the same metrics. Still, each study is different in choices of criteria, data splits, cohort variables, time duration, the metric used, and other factors. This work has observed three primary metrics for measuring algorithm prediction accuracy: Harrell's concordance index (C-index), Integrated Barrier score, and the Area Under the Curve(AUC). The C-index is not used for neural network techniques because subjects can not be naturally ordered. Some papers made an indirect comparison of ML algorithms to the Cox model to examine the most prognostic variables.

Table 2: Results of Algorithms compared .

Paper	ML	ML‖OM	Other
7	`SVM`	.800 ‖ .620	LR
4	`RF`	.718 ‖ .706	Cox
6	`ANN`	.764 ‖ .649	LR
8	`SDT`	.800 ‖ .710	CDT
5	`GB`	.644 ‖ .627	Coxnet
6	`DNN`	.691 ‖ .652	AdaBoost

OM: Other Method
SDT: Survival Decision Tree
CDT: Conventional Decision Tree
GB: Gradient Boosting

5 Discussion

The outcome of organs transplant prediction for decision-making is vital in clinical settings. The addition of more relevant features can increase the accuracy of a prediction system. The study shows that there could be a save of $750 million if transplant prediction can be improved by .057. Apart from the forecast, ML techniques have been used to discover patterns and data exploration, which helped choose the appropriate algorithm. The stability of these tools is guaranteed based on several runs performed with the same parameter. This study has helped to realize that machine learning interpretation is feasible for different variable levels with the method of Garson. This review study shows that most survival analyses are either a classification problem or a survival function modeling. The limitation of this review includes a few numbers of selected papers for the review as a result of the time frame to do more depth search and analyses.

6 Conclusion

This review shows that machine learning techniques would be more beneficial for both predictions of organ transplant and interpretation of the result. This review is helpful to clinical decision-making to encourage machine learning algorithms for the prediction process in organ transplant systems. The future recommendation is to develop metrics that directly interpret predictions to clinicians and patients to avoid the so-called black-box challenge of interpretability.

References

[1] Kantidakis G.;Putter H.;Lancia C.;Boer J.;Braat A.E. and Fiocco M. Survival prediction models since liver transplantation - comparisons between cox models and machine learning techniques. *BMC Medical Research Methodology*, 2020. https://doi.org/10.1186/s12874-020-01153-1.

[2] Pahl E.S.;Street W.N.;Johnson H.J.; Reed A.I. A predictive model for kidney transplant graft survival using machine learning. *International Conference on Computer Science and Information Technology*, 2020. https://arxiv.org/abs/2012.03787.

[3] Biganzoli E.;Borracchi P.; Mariani L.; Marubini E. Feed forward neural network for the analysis of censored survival data: a partial logistic regression approach. 1998 journal =.

[4] Mark E.;Goldsman D.;Gurbaxani B.;Keskinocak P.;Sokol J. Using machine learning and an ensemble of methods to predict kidney transplant survival. *Creative Commons CCO*, 2019. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0209068.

[5] Nemati M.;Zhang H.;Sloma M.;Bekbolsynov D.;Wang H.;Stepkowski S.;Xu K.S. Predicting kidney transplant survival using multiple feature representations for hlas. 2021. https://arxiv.org/abs/2103.03305.

[6] Ayers B.; Sandholm T.; Gosev I.; Prasad S. and Kilic A. Using machine learning to improve survival prediction after heart transplantation. *Journal of Cadiac Surgery*, 2021. wileyonlinelibrary.com/journal/jocs.

[7] Naqvi S.A.;Tennankore K.; Vinson A.; Roy P.C.; Abidi S.S. Predicting kidney graft survival using machine learning methods prediction model development and feature significance analysis

study. *Journal of Medical Internet Research*, 2021. https://pubmed.ncbi.nlm.nih.gov/34448704/.

[8] Yoo K.D.; Noh J.; Lee H.; Kim D.K.; Lim C.S.; Kim Y.H.; Lee J.P.; Kim G.; Kim Y.S. A machine learning approach using survival statistics to predict graft survival in kidney transplant recipients: A multicenter cohort study. *National Librbary of Medicine*, 2017. https://pubmed.ncbi.nlm.nih.gov/28827646/.

[9] Killian M.O.;Payrovnaziri S.N.;Gupta D.;Desai D.;He Z. Machine learning–based prediction of health outcomes in pediatric organ transplantation recipients. *JAMIA Open*, 2020. https://academic.oup.com/jamiaopen/article/4/1/ooab008/6168494.