Machine-learning-assisted donor-recipient matching for orthotopic liver transplantation

Bezjak, Miran; Stresec, Ivan; Jadrijević, Stipislav; Kocman, Branislav; Filipec Kanižaj, Tajana; Antonijević, Miro; Dalbelo Bašić, Bojana; Mikulić, Danko

Conference presentation / Izlaganje na skupu

https://doi.org/10.1097/01.tp.0000978836.44371.fe

ILTS ANNUAL CONGRESS 2023
Rotterdam, Netherlands
MAY 3-6, 2023

ELITA ESOT

Machine-Learning-Assisted Donor-Recipient Matching for Orthotopic Liver Transplantation

M. BEZJAK [1,*], I. STRESEC [2,*], S. JADRIJEVIĆ [1], B. KOCMAN [1], T. FILIPEC KANIŽAJ [3], M. ANTONIJEVIĆ [4], B. DALBELO BAŠIĆ [2], D. MIKULIĆ [1]

1 University Hospital Merkur, Department of Surgery, Zagreb, Croatia
2 University of Zagreb, Faculty of Electrical Engineering and Computing, Zagreb, Croatia
3 University Hospital Merkur, Department of Internal Medicine, Zagreb, Croatia
4 KONČAR Digital, Zagreb, Croatia
*These authors contributed equally.

INTRODUCTION

Liver transplant allocation policies evolve over time. With no universal algorithm to predict the outcome of liver transplantation allocation, donor-recipient matching still relies heavily on the experience of the transplant team of the institution. **Machine learning (ML) models, using data collected on-site, could offer more reliable and relevant ranking systems in comparison to more traditional prognostic scores.**

AIM

In this work, we test several survival ML algorithms against the BAR [1] and ET-DRI [2] scores to see if a localized ML approach provides a more accurate metric for survival prediction, potentially mitigating locational dataset shifts while also leveraging all available variables. The event of interest is overall graft failure, i.e., failure-free survival. **We use a robust testing framework to obtain realistic assessments of the concordance index** (c-index) of our ML models' performances. We interpret variable importance within the framework.

METHODS

24 donor and recipient variables were collected from 656 patients who underwent liver transplantation from March 2013 through December 2018 at the University Hospital Merkur (Table 1). **We developed a robust evaluation system** and tested several different models: regularized Cox regression (CoxNet) [3] as a linear baseline; Random Survival Forest (RSF) [4]; gradient-boosted trees (GBT) [5]; and Survival Support Vector Machine (SSVM) [6]. The **models were evaluated using 5 times repeated nested cross-validation** (RNCV), with 10 folds for both outer and inner CV. Hyperparameters were selected in grid searches. Before fitting the models, categorical variables were one-hot encoded, missing values were imputed using iterative imputation [7], and all variables were standardized, in that order. Imputation and standardization were fitted on training data within the RNCV procedure, avoiding potential bias. The **models were interpreted using Shapley additive explanations (SHAP)** [8]. SHAP values are averaged across the test sets within the RNCV procedure [9]. As a measure of comparison between models (in all steps of the procedure), **we employed Uno's c-index estimator [10], appropriate for the high censorship rates in our data (72%)**.

RESULTS

Our best ranking system, based on GBT, achieved an average c-index of 64.8%, outperforming traditional BAR and ET-DRI scores that achieve a concordance index of only 52.2% and 53.2%, respectively. RSF performed similarly to GBT, while the CoxNet and SSVM models showed inferior ranking capability. The scores are shown in detail in Table 2 and Figure 1. GBT and RSF model nonlinearities and interactions, which is likely why they performed so well. CoxNet is linear and does not model nonlinearities or interaction, making its relatively low c-index expected. The SSVM model can generally handle complex data and a wider grid search could potentially improve its performance, but it was left out due to costly training.

Interpretation via SHAP values yielded many relevant predictors of survival: recipient and donor age and BMI, donor weight, MELD, donor sodium (Na), donor CRP, donor ALT, donor GGT, CIT, and recipient diagnoses (Dx). The importance of many variables indicates good utilization of available data. Additionally, steatosis was deemed the most important variable in terms of SHAP values, but its actual importance is disputable due to many missing values. The SHAP interpretation can be seen in Figure 2.

	c-index
BAR	52.2% (–)
ET-DRI	53.2% (–)
CoxNet	55.0% (± 4.5%)
SSVM	54.8% (± 4.2%)
RSF	62.8% (± 5.1%)
GBT	64.8% (± 5.3%)

Table 2: The c-index and its standard deviation for each model, estimated on test folds. The c-indexes for BAR and ET-DRI are calculated on the entire dataset. All c-indexes are estimated using Uno's estimator.

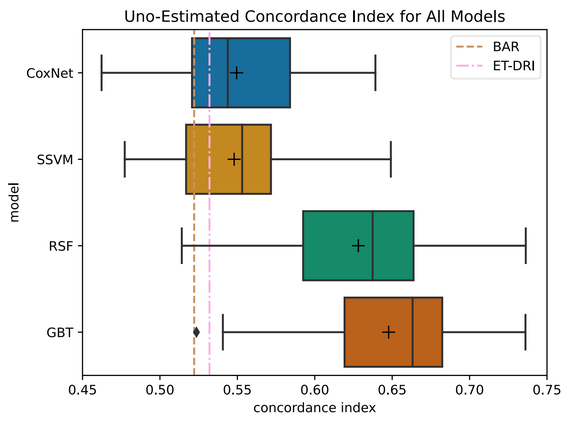

Figure 1: A side-by-side boxplot of the c-index of all tested models. The c-indexes of the BAR and ET-DRI scores are drawn with a dashed and dashed-dotted line, respectively, for visual comparison. Plus (+) represents the mean value. The relatively large variance in model performance can partly be attributed to small test sets.

Table 1: A short summary of the University Hospital Merkur dataset.

	1 year	3 years	5 years	Overall
Survival Rate	74.8%	55.2%	20.3%	–
Censorship Rate	15.9%	39.5%	64.6%	71.6%
Survival Median	629 days			

Donor variables	age, weight, height, BMI, sex, blood type, anti-HBc, steatosis, sodium, CRP, ALT, GGT, bilirubin, CIT, cardiac arrest, pancreas explantation offer
Recipient variables	age, weight, height, BMI, sex, blood type, MELD, diagnosis
Missing values	steatosis (125), CRP (46), bilirubin (24), GGT (8), ALT (2), sodium (2), CIT (2)

Figure 2: A beeswarm summary plot of the SHAP values of the top 10 influential variables for the GBT model. Positive SHAP values imply a negative impact on survival. The SHAP values of each test sample are averaged across outer folds and repeats of the RNCV procedure. The missing values of steatosis are shown in grey. RSF has a near-identical SHAP plot, implying relatively consistent variable importance across well-performing models. AC is short for alcoholic cirrhosis.

CONCLUSIONS

With ML, we can improve upon existing metrics by creating an interpretable ranking system that better fits available data. Such systems can provide superior assistance in donor-recipient matching while also identifying relevant risk factors. **The benefits of using ML models trained on local data in comparison to more traditional risk scores are the potential mitigation of locational dataset shifts and leveraging of all available variables.** However, this requires that a hospital gathers significant amounts of data, and the interpretation of models is less straightforward. Moreover, the c-index, while most often used in survival ML literature, is not without flaws. We aim to address these issues in future work.

REFERENCES

[1] **Dutkowski P, Oberkofler CE, Slankamenac K, Puhan MA, Schadde E, Müllhaupt B, et al.** Are there better guidelines for allocation in liver transplantation? A novel score targeting justice and utility in the model for end-stage liver disease era. Annals of Surgery. 2011 Nov;25 (5):745-53; discussion 753. Available from: https://doi.org/10.1097/SLA.0b013e3182365081.

[2] **Braat AE, Blok JJ, Putter H, Adam R, Burroughs AK, Rahmel AO, et al.** The Eurotransplant donor risk index in liver transplantation: ET-DRI. American Journal of Transplantation: Official Journal of the American Society of Transplantation and the American Society of Transplant Surgeons. 2012 Oct;12(10):2789-96. Available from: https://doi.org/10.1111/j.1600-6143.2012.04195.x.

[3] **Simon N, Friedman J, Hastie T, Tibshirani R.** Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. Journal of Statistical Software. 2011 Mar;39(5):1-13. Available from: https://doi.org/10.18637/jss.v039.i05.

[4] **Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS.** Random survival forests. The Annals of Applied Statistics. 2008;2(3):841 860. Available from: https://doi.org/10.1214/08-AOAS169.

[5] **Hothorn T, Bühlmann P, Dudoit S, Molinaro A, Van Der Laan MJ.** Survival ensembles. Biostatistics. 2006 Jul;7(3):355-73. Available from: https://doi.org/10.1093/biostatistics/kxj011.

[6] **Pölsterl S, Navab N, Katouzian A.** An Efficient Training Algorithm for Kernel Survival Support Vector Machines. arXiv; 2016. ArXiv:1611.07054 [cs, stat]. Available from: http://arxiv.org/abs/1611.07054.

[7] **van Buuren S, Groothuis-Oudshoorn K.** mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software. 2011 Dec;45(3):1-67. Available from: https://www.jstatsoft.org/index.php/jss/article/view/v045i03.

[8] **Lundberg S, Lee SI.** A Unified Approach to Interpreting Model Predictions. arXiv; 2017. ArXiv:1705.07874 [cs, stat]. Available from: http://arxiv.org/abs/1705.07874.

[9] **Scheda R, Diciotti S.** Explanations of Machine Learning Models in Repeated Nested Cross-Validation: An Application in Age Prediction Using Brain Complexity Features. Applied Sciences. 2022 Jul;12(13):6681. Available from: http://dx.doi.org/10.3390/app12136681.

[10] **Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ.** On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. Statistics in Medicine. 2011;30(10):1105-17. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.4154.

ACKNOWLEDGEMENTS

This research has been supported by the European Regional Development Fund under the grant KK.01.1.1.01.0009 (DATACROSS).

CONTACT INFORMATION

miran.bezjak@gmail.com