ESC HEART FAILURE
*ESC Heart Failure* 2021; **8**: 3433–3436
Published online 18 May 2021 in Wiley Online Library (wileyonlinelibrary.com) **DOI:** 10.1002/ehf2.13425

LETTER TO THE EDITOR

# State-of-the-art machine learning improves predictive accuracy of 1-year survival after heart transplantation

Heart transplantation (HT) remains the treatment of choice for patients with medically refractory end-stage heart failure given its improved long-term outcomes and quality of life.[1] Despite the effectiveness of the treatment, only about 3000 HTs are performed annually in the USA, with a small rise attributed to the opioid epidemic and the expanded use of donors with hepatitis C.[2] Optimization of outcomes and risk stratification is recognized as a critically important issue in HT today.[3] Although several risk scores have been developed for the prediction of outcomes after HT, their accuracy remains modest. Machine learning (ML) algorithms have shown promising results in predicting outcomes and discovering phenotypes in patients with heart failure[4] and have even been applied in patients undergoing HT.[5] However, further work is needed to determine the role of ML in the prediction of outcomes and risk stratification of HT patients. Herein, we present our work on prediction of 1-year survival after HT using state-of-the-art ML algorithms.

We queried the United Network for Organ Sharing (UNOS) database for patients that were enlisted and underwent heart transplant between 2010 and 2018. Extensive baseline clinical and laboratory data as well as follow-up survival were collected. We excluded patients with multiple transplants during the index hospitalization, non-adult patients, and patients with >10% missing values or less than 1-year follow-up. The study dataset was then randomly split into training and validation cohorts with a 3:1 ratio. A random oversampling technique was applied to artificially equalize the number of patients that were alive with the number of patients that did not survive at 1-year in the training cohort. This is common practice for many ML algorithms. Next, a feature selection method based on Support Vector Machines (SVM) and Fuzzy Logic, namely, SVM-FuzCoC,[6] was used to (i) identify the most informative variables, (ii) reduce the complexity, and (iii) improve the performance of the ML models. The selected features were then used to train five state-of-the-art supervised ML algorithms: Adaptive boosting (Adaboost), SVM, Decision Trees, K-nearest neighbour, and Logistic Regression.[7] Training of the models was controlled by hyperparameter optimization. After training, the validation cohort was used to internally test the ML algorithms. Their performance was evaluated by measuring the total area under the curve (AUC) of the receiver-operator curve as well as sensitivity, specificity, and positive and negative predictive value. We performed explainability analysis of the best ML model using the method of local interpretable model-agnostic explanations (LIME).[8] This method creates a simple linear model that approximates any complex ML model. We separately applied the IMPACT score for the prediction of 1-year survival after HT.

After preprocessing, the dataset composed of 18 625 recipients (mean age 53 ± 13 years, 73% male) with 134 pre-transplant variables. There were no significant differences between patients assigned to the development and validation cohort. There were a total of 2334 (12.5%) deaths at 1-year after HT. Feature selection identified 39 out of 134 variables that were highly predictive of 1-year survival and were used in the ML algorithms (*Table 1*). These included recipient, donor, and transplant characteristics. AUC for each of the five ML using the validation cohort is shown in *Figure 1A*. Although the best performance was achieved by Adaboost (AUC 0.689, 95% CI 0.665–0.715), all ML models achieved higher AUC compared with IMPACT score with the exception of the K-nearest neighbour (SVM AUC 0.637 95% CI 0.612–0.662, DT AUC 0.649 95% CI 0.622–0.676, LR AUC 0.642 95% CI 0.614–0.667, and KNN AUC 0.527 95% CI 0.502–0.550 in contrast to IMPACT AUC 0.569, 95% CI 0.545–0.592). With the exception of K-nearest neighbour, there was an exponential increase in AUC with the use of one to five pre-transplant variables. However, the increase in AUC gradually plateaued after the use of 10 variables. Sensitivity, specificity, and positive and negative predictive value of the ML models using the validation subcohort ranged from 55.6% to 68.5%, 42.6% to 63.4%, 88.7% to 93.6%, and 13.5% to 21.6%, respectively. The results of explainability analysis for the best ML model using LIME are shown in *Figure 1B*. Increased serum creatinine, shorter recipient height, and increased ischaemic time had the strongest association with 1-year mortality.

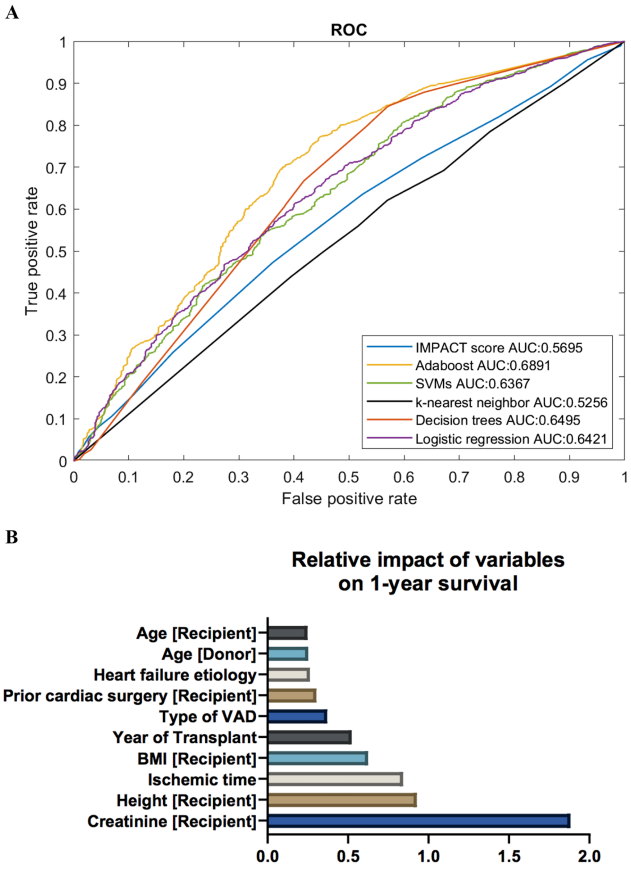
Artificial intelligence and ML algorithms in particular offer some striking advantages compared with traditional statistical methods in regard to their ability to analyse large and

**Table 1** Baseline donor, recipient, and transplant characteristics selected by the feature selection algorithm as highly relevant to 1-year survival

| | Demographic | History/clinical | Laboratory |
|---|---|---|---|
| **Recipient** | Age (years) | Ventricular assist device | Creatinine (mg/dL) |
| | Sex | Prior cardiac surgery | HIV seropositive |
| | Ethnicity | Heart failure etiology | EBV seropositive |
| | Height (cm) | Waiting time in 1A status (days) | CMV seropositive |
| | BMI (kg/m$^2$) | Total waiting time (days) | HBV (HBsAg+) |
| | Level of education | Treatment with prostaglandins | |
| | | AICD | |
| | | Inotropic support | |
| | | Mechanical ventilation | |
| **Donor** | Age (years) | Alcohol abuse | Creatinine (mg/dL) |
| | Ethnicity | Cocaine abuse | SGPT (mg/dL) |
| | BMI (kg/m$^2$) | Inotropic support | $PCO_2$ (mmHg) |
| | | HTN requiring medical therapy | |
| | | Cause of death | |
| | | **Characteristic** | |
| **Transplant** | | Ischaemic time (min) | |
| | | Transplant surgery duration (min) | |
| | | Transplant year[a] | |
| | | Listing year | |
| | | ABO match | |

BMI, body mass index.
These variables were used to train the machine learning algorithms. Continuous variables have their units in parentheses.
[a]Within 2010–18.

**Figure 1** (A) Receiver-operating characteristic curves for ML models and the IMPACT score. (B) LIME analysis visualizing relative impact of top-10 variables on 1-year survival for the best performing ML model (Adaboost). Note that smaller recipient height was associated with decreased survival. Results in logarithmic scale. AUC, area under the curve; ML, machine learning; SVM, support vector machine.





DOI: 10.1002/ehf2.13425

multivariable datasets. The previously published International Heart Transplant Survival Algorithm (IHTSA) was based on a flexible, non-linear artificial neural network to predict survival after HT, included over 56 000 adult patients who underwent HT from 1994 to 2010 with a reported AUC 0.65 for 1-year survival.[5] Our analysis included HT recipients after 2010 including patients transplanted after the recent change in the allocation policies, thus representing contemporary practices and outcomes. In addition, we used a feature selection technique, which identified predictors of 1-year mortality and LIME explainability analysis. On the other hand, the IMPACT score was based on patients that received HT before 2010, and an era affect may partially explain its low performance in our analysis.

Predictors of mortality from our analysis are in accordance with recently published big data analyses.[9] Increased pre-transplant serum creatinine in particular is a well-known predictor of mortality that has been strongly associated with mortality after HT[10] and was found to have the largest impact on outcomes in our LIME explainability analysis of the Adaboost ML model. Although this finding does not add clinical novelty by itself, it represents a novel way to confirm that the ML model is based on clinically solid grounds and can therefore be used to predict outcomes. Shorter stature also strongly impacted 1-year survival in the Adaboost ML model. Although we cannot entirely explain this finding, we believe it is related to height mismatch, which has been associated with poor outcomes after HT in the past. A focused analysis on height mismatch was not a primary goal of this analysis. Our results additionally draw attention to the potential importance of donor parameters such as history of alcohol and cocaine use, cause of death, and renal and liver indices, at least as hypothesis generating.

In conclusion, ML models created and validated using a contemporary cohort of the UNOS database showed improved accuracy in predicting 1-year survival after HT. Further validation of these models in other HT cohorts and the assessment of their predictive capacity of long-term survival after HT are warranted.

## Acknowledgement

The authors thank Dr Gerasimos Filippatos for his helpful input.

## Funding

None.

Polydoros N. Kampaktsis
*Division of Cardiology, New York University Langone Medical Center, New York City, NY, USA*

Serafeim Moustakidis
*AiDEAS, Tallinn, Estonia*

Aspasia Tzani
*Brigham and Women's Hospital Heart and Vascular Center, Harvard Medical School, Boston, MA, USA*

Ilias P. Doulamis
*Division of Cardiac Surgery, Boston's Children Hospital, Boston, MA, USA*

Anastasios Drosou
*Information Technologies Institute, National Center for Research and Technology, Thessaloniki, Greece*

Andreas Tzoumas
*Aristotle University of Thessaloniki Medical School, Thessaloniki, Greece*

Rabea Asleh
*Hadassah Medical Center, Jerusalem, Israel*

Alexandros Briasoulis
*Division of Cardiovascular Medicine, University of Iowa Carver College of Medicine, Iowa City, IA, USA*
*E-mail: alexbriasoulis@gmail.com*

## References

1. Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE Jr, Drazner MH, Fonarow GC, Geraci SA, Horwich T, Januzzi JL, Johnson MR, Kasper EK, Levy WC, Masoudi FA, McBride PE, McMurray JJ, Mitchell JE, Peterson PN, Riegel B, Sam F, Stevenson LW, Tang WH, Tsai EJ, Wilkoff BL. 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol* 2013; **62**: e147–e239.
2. Mehra MR, Jarcho JA, Cherikh W, Vaduganathan M, Lehman RR, Smits J, Stehlik J. The drug-intoxication epidemic and solid-organ transplantation. *N Engl J Med* 2018; **378**: 1943–1945.
3. Kittleson MM, Kobashigawa JA. Cardiac transplantation: current outcomes and contemporary controversies. *JACC: Heart Fail* 2017; **5**: 857–868.
4. Jing L, Cerna AEU, Good CW, Sauers NM, Schneider G, Hartzel DN, Leader JB, Kirchner HL, Hu Y, Riviello DM, Stough JV, Gazes S, Haggerty A, Raghunath S, Carry BJ, Haggerty CM, Fornwalt BK. A machine learning approach to management of heart failure populations. *JACC: Heart Fail* 2020; **8**: 578–587.
5. Nilsson J, Ohlsson M, Höglund P, Ekmehag B, Koul B, Andersson B. The International Heart Transplant Survival Algorithm (IHTSA): a new model to improve organ sharing and survival. *PLoS ONE* 2015; **10**: e0118644.
6. Moustakidis SP, Theocharis JB. SVM-FuzCoC: a novel SVM-based feature selection method using a fuzzy

complementary criterion. *Pattern Recogn* 2010; **43**: 3712–3729.
7. Witten IH, Frank E, Hall MA. *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed. Burlington, MA: Morgan Kaufmann; 2011.
8. Ribeiro MT, Singh S, Guestrin C. Why should I trust you? Explaining the predictions of each classifier. KDD; 2016; San Francisco, CA2016.
9. Hsich EM, Blackstone EH, Thuita LW, McNamara DM, Rogers JG, Yancy CW, Goldberg LR, Valapour M, Xu G, Ishwaran H. Heart transplantation. *JACC: Heart Fail* 2020; **8**: 557–568.
10. Odim J, Wheat J, Laks H, Kobashigawa J, Gjertson D, Osugi A, Mukherjee K, Saleh S. Peri-operative renal function and outcome after orthotopic heart transplantation. *J Heart Lung Transplant* 2006; **25**: 162–166.


DOI: 10.1002/ehf2.13425