Using Machine Learning to Improve Survival Prediction After Heart Transplantation

Brian Ayers[1], Tuomas Sandholm[2], Igor Gosev[3], Sunil Prasad[3], and Arman Kilic[4]

[1]Massachusetts General Hospital
[2]Carnegie Mellon University
[3]University of Rochester Medical Center
[4]Medical University of South Carolina

April 28, 2021

Abstract

Background: This study investigates the use of modern machine learning (ML) techniques to improve prediction of survival after orthotopic heart transplantation (OHT). Methods: Retrospective study of adult patients undergoing primary, isolated OHT between 2000-2019 as identified in the United Network for Organ Sharing (UNOS) registry. The primary outcome was one-year post-transplant survival. Patients were randomly divided into training (80%) and validation (20%) sets. Dimensionality reduction and data re-sampling were employed during training. Multiple machine learning algorithms were combined into a final ensemble ML model. Discriminatory capability was assessed using area under receiver-operating-characteristic curve (AUROC), net reclassification index (NRI), and decision curve analysis (DCA). Results: A total of 33,657 OHT patients were evaluated. One-year mortality was 11% (n=3,738). In the validation cohort, the AUROC of singular logistic regression was 0.649 (95% CI 0.628-0.670) compared to 0.691 (95% CI 0.671-0.711) with random forest, 0.691 (95% CI 0.671-0.712) with deep neural network, and 0.653 (95% CI 0.632-0.674) with Adaboost. A final ensemble ML model was created that demonstrated the greatest improvement in AUROC: 0.764 (95% CI 0.745-0.782) ($p<0.001$). The ensemble ML model improved predictive performance by 72.9% ±3.8% ($p<0.001$) as assessed by NRI compared to logistic regression. DCA showed the final ensemble method improved risk prediction across the entire spectrum of predicted risk as compared to all other models ($p<0.001$). Conclusions: Modern ML techniques can improve risk prediction in OHT compared to traditional approaches. This may have important implications in patient selection, programmatic evaluation, allocation policy, and patient counseling and prognostication.

Using Machine Learning to Improve Survival Prediction After Heart Transplantation

Authors: Brian Ayers MD MBA[1], Tuomas Sandholm PhD[2], Igor Gosev MD PhD[3], Sunil Prasad MD[3], Arman Kilic MD[4]

Affiliations:

[1] Department of Surgery, The Massachusetts General Hospital, Boston, MA

[2] Carnegie Mellon University, Pittsburgh, PA

[3] Division of Cardiac Surgery, The University of Rochester Medical Center, Rochester, NY

[4] Division of Cardiothoracic Surgery, Medical University of South Carolina, Charleston, SC

Short Title: Machine Learning Heart Transplant

Word Count: 3,516

Subject Terms: heart failure; transplantation; complications; mortality/survival

Meeting Presentation: American Heart Association Scientific Sessions 2020 (virtual meeting)

Funding: None

Address for correspondence:

Arman Kilic, M.D.

Division of Cardiothoracic Surgery

Medical University of South Carolina

30 Courtenay Drive

MSC 295, Suite BM279

Charleston, SC 29425

Email: kilica@musc.edu

Abstract:

Background: This study investigates the use of modern machine learning (ML) techniques to improve prediction of survival after orthotopic heart transplantation (OHT).

Methods: Retrospective study of adult patients undergoing primary, isolated OHT between 2000-2019 as identified in the United Network for Organ Sharing (UNOS) registry. The primary outcome was one-year post-transplant survival. Patients were randomly divided into training (80%) and validation (20%) sets. Dimensionality reduction and data re-sampling were employed during training. Multiple machine learning algorithms were combined into a final ensemble ML model. Discriminatory capability was assessed using area under receiver-operating-characteristic curve (AUROC), net reclassification index (NRI), and decision curve analysis (DCA).

Results: A total of 33,657 OHT patients were evaluated. One-year mortality was 11% (n=3,738). In the validation cohort, the AUROC of singular logistic regression was 0.649 (95% CI 0.628-0.670) compared to 0.691 (95% CI 0.671-0.711) with random forest, 0.691 (95% CI 0.671-0.712) with deep neural network, and 0.653 (95% CI 0.632-0.674) with Adaboost. A final ensemble ML model was created that demonstrated the greatest improvement in AUROC: 0.764 (95% CI 0.745-0.782) ($p<0.001$). The ensemble ML model improved predictive performance by 72.9% ±3.8% ($p<0.001$) as assessed by NRI compared to logistic regression. DCA showed the final ensemble method improved risk prediction across the entire spectrum of predicted risk as compared to all other models ($p<0.001$).

Conclusions: Modern ML techniques can improve risk prediction in OHT compared to traditional approaches. This may have important implications in patient selection, programmatic evaluation, allocation policy, and patient counseling and prognostication.

Words: 248

Non-Standard Abbreviations and Acronyms

AUROC: area under receiver-operating-characteristic curve

DCA: decision curve analysis

DRI: donor risk index

IHTSA: international heart transplant survival algorithm

IMPACT: index for mortality prediction after cardiac transplantation

ML: machine learning

NRI: net reclassification index

OHT: orthotopic heart transplantation

RSS: risk stratification score

STS: Society of Thoracic Surgeons

UNOS: United Network for Organ Sharing

Introduction

The clinical management of end-stage heart disease is a continually evolving practice that has changed drastically over the last decade. Survival after orthotopic heart transplantation (OHT) has continued to improve.[1] Despite improving longevity after transplant, donor organ supply remains inadequate to meet demand,[2] and transplant programs face increased public and private scrutiny of their outcomes.[3]Simultaneously, technologic innovations in mechanical circulatory support platforms have demonstrated parallel improvement in clinical outcomes,[4-6] thus increasing the potential alternative viable treatment options for heart failure patients. Therefore, an accurate prognostic model using pre-operative data for individualized donor and recipient selection would be of profound clinical utility in OHT.

Prior risk models for predicting survival after OHT have displayed only modest discriminatory capability. Examples of such algorithms include The Donor Risk Index (DRI),[7] Risk Stratification Score (RSS),[8] Index for Mortality Prediction After Cardiac Transplantation (IMPACT)[9] and International Heart Transplant Survival Algorithm (IHTSA)[10]. With increasing interest in the application of machine learning (ML) to predictive analytics in clinical medicine,[11] we aimed to evaluate whether modern ML techniques could improve risk prediction in OHT.

Material and Methods

Patient Population

Because of the sensitive nature of the data collected for this study, requests to access the dataset from qualified researchers trained in human subject confidentiality protocols may be sent to the United Network for Organ Sharing (UNOS). We conducted a retrospective review of prospectively collected data in the UNOS database. The database was queried for all patients that underwent orthotopic heart transplantation (OHT) between 2000-2019. Patients were excluded if they underwent transplantation before the year 2000, were less than 18 years old, had a history of prior heart transplantation, underwent multiorgan transplant, or had incomplete survival status at one year. This retrospective analysis of deidentified data was deemed except from Institutional Review Board approval and patient consent was not required.

Training and Validation Cohorts

Patients were randomly split into training (80%) and validation (20%) cohorts, ensuring equal distribution of the primary outcome in each cohort. This method of stratifying by primary outcome before randomly assigning patients to a cohort ensures an equal distribution of mortality between training and validation datasets to avoid biasing final model performance. The training cohort was used for feature selection, dimensionality reduction, and machine learning model development, keeping the validation cohort entirely separate and unseen until assessment of the final model performance.

Data Preparation and Feature Selection

All variables in the UNOS database available for the OHT patients were manually reviewed by two independent clinicians (N=525 variables). Variables were excluded if they were redundant, free text, or would not be available in the preoperative setting. Variables with more than 20% missing data were also excluded. The distribution of data for each remaining categorical variable were again manually reviewed and grouped into clinically meaningful categories for each variable by two independent physicians. This step decreases data sparsity by grouping low incidence characteristics into fewer, clinically meaningful categories. Missing continuous variable data were imputed using feature median and missing categorical data were imputed with

the feature mode. Continuous variables were standardized to have a mean of zero and standard deviation of one. Categorical variables were one-hot encoded ensuring no linear dependencies between columns.

Feature Importance and Dimensionality Reduction

In order to increase the interpretability of the model, we first performed a univariate estimate of feature importance. A random forest classifier was trained with 10-fold cross-validation repeated three times. Univariate feature importance was estimated using the mean decrease in accuracy. Three separate feature importance models were developed first with continuous variables only (N=37), then with categorical variables (N=195), and then with all the variables (N=232). This was done to account for the known tendency for random forest classifiers to more heavily weight continuous variables in feature importance estimates. The 20 most important features from each model were combined into a single dataset, consisting of 47 variables after excluding duplicates. This variable set was subsequently used to develop the machine learning algorithms. Overall, this methodology of feature selection allows us to combine the best of both manual filtering based on clinical acumen and automated methods using machine learning techniques. Decreasing the input space used for the final machine learning model is a critical step that has been shown to improve overall model performance by decreasing tendency to overfit and increasing training efficiency.[12]

Training Prediction Models

An ensemble approach was employed in order to create a more stable and reliable model resistant to outliers. Four different types of algorithms were used: deep neural network, logistic regression, adaboost, and random forest. We employed over-sampling (SVM Smote) and under-sampling (Repeated Edited Nearest Neighbors) of the training data in order to better balance the primary outcome within the dataset. For each type of algorithm, 100 different models were trained using varying degrees of data re-sampling to produce variability in each model's underlying training data. The resulting 400 algorithms were subsequently combined into the final ensemble prognostic model.

Validation and Comparison

Once the models were trained, discriminatory capability was assessed using the previously unseen validation data. The performance of the full ensemble model (400 algorithms) was compared to that of each type of algorithm individually (100 algorithms each), as well as a single logistic regression on its own. Model capability was assessed using area under receiver-operating-characteristic curve (AUROC), net reclassification index (NRI), and decision curve analysis (DCA). Calibration of the model was evaluated using visual plots of predicted risk based on the training cohort versus observed risk in the validation cohort stratified by decile of risk. Two-sided p-value of less than 0.05 was considered significant for all comparisons. All models were trained in python using Keras with Tensorflow.[13] Performance outcome comparisons were conducted with Stata (StataCorp. 2015. Stata Statistical Software: Release 14. College Station, TX: StataCorp LP).

Results

Study Cohort

Of the 33,657 patients included in the study, 3,728 (11%) experienced the primary outcome of death within one year (**Figure 1**). The majority of study participants were male (75%) with an average age of 52.8 ±12.4 years. Balancing by primary outcome, patients were randomly split into 80% training cohort and 20% validation cohort. Patient characteristics were similar between cohorts, including age, sex, etiology of heart failure, end-organ function, pre-transplant mechanical circulatory support, and days spent on waiting list (**Table 1**).

Feature Importance and Dimensionality Reduction

After manual and automated variable filtering, the relative importance of each feature in predicting the primary outcome was assessed using a random forest classifier. Continuous and categorical variables were assessed both separately and together to reduce biases (**Supplemental Figure 1**). In the combined

assessment, the most important features included total bilirubin, mechanical ventilation, no ventricular assist device at transplant, serum creatinine, donor age, recipient height, donor cerebrovascular mechanism of death, prior cardiac surgery, and Karnofsky functional status. The top 20 most important features from each individual assessment were then combined excluding duplicates, resulting in 47 training variables that were used for subsequent model training (**Supplemental Table 1**).

Model Training and Performance Comparison

A total of 400 algorithms were trained using varying subsets of training data based on randomly stratified levels of over and under resampling of the training dataset (**Figure 2**). The final ensemble ML model contained all 400 underlying algorithms, while smaller sized ensemble models were also combined using the 100 iterations of each type of algorithm individually for comparison. The optimal model performance was the complete ensemble ML model (**Figure 3**), outperforming all other models with an AUROC of 0.764 (95% CI, 0.745-0.782) ($p<0.001$). By comparison, the singular logistic regression model had an AUROC of 0.649 (95% CI, 0.628-0.670). Additionally, the final ensemble ML model demonstrated an improvement of 72.9% $\pm$3.8% ($p<0.001$) in predictive performance as assessed by net reclassification index compared to logistic regression. The decision curve analysis showed the final ensemble method improved risk prediction across the entire spectrum of predicted risk as compared to all other models (**Figure 4** , $p<0.001$). The final ensemble ML model was well-calibrated, with the majority of observed risk in the validation cohort falling within range of predicted risk based on the training cohort after stratifying into deciles of risk (**Supplemental Figure 2**).

Conclusions

Prognostication of clinical outcomes after OHT has profound importance in patient selection and organ allocation. The present study demonstrates the potential utility of employing modern ML techniques to improve prognostic model performance at an individual patient level. A final ensemble ML model using only preoperative variables outperformed all other comparison algorithms in predicting one-year survival, achieving improved performance by a variety of metrics including AUROC, net reclassification index, and decision curve analysis. Further, the model demonstrated appropriate calibration.

Cardiac surgery as a field has historically been an early adopter of clinical prognostic models,[14] most notably the widely used Society of Thoracic Surgeons (STS) Short-Term Risk Calculators.[15] However, predicting one-year mortality after OHT has remained a persistent challenge. Early OHT risk models incorporated a select number of variables with only modest overall performance.[7,8,16] More recent models have added an increasing number of variables into more robust models, such as the IMPACT[9] and IHTSA[10], but have been able to achieve only slight improvements in discriminatory performance. In part, this may relate to the challenges in capturing all granular and potentially predictive elements of post-transplant survival in a multicenter registry. For example, factors such as anti-rejection medication compliance are not assessed but can have important implications in survival following transplant. Also, there is a trade-off in assessing longer term outcomes, such that the event rate will be higher but the impact of pre-operative risk factors on that outcome will likely diminish as longer-term factors weight more heavily into outcome prediction.

Machine learning techniques have demonstrated clinical utility in a number of different fields.[17-21] Within OHT, the IHTSA score itself employs an artificial neural network approach, and has consistently demonstrated some of the highest discriminatory values out of all current models in recent studies.[22] Moreover, a recent study recalibrated both the IMPACT and IHTSA models to use only the same subset of variables, and found the deep learning approach was superior.[23] The recent Trees of Predictors model is also an innovative approach that identifies clusters of patients with similar characteristics, and develops machine learning predictive models specifically for each cluster.[24] The success of this approach demonstrates the potential for developing very individualized prognostic scores, at the risk of overfitting the model to specific retrospective cohorts that may not translate to prospective clinical practice.

The final ensemble ML model we developed in the present study is an example of using both clinical acumen and automated machine learning to develop a robust model from a large clinical registry. The statistical adage of "garbage in produces garbage out" remains especially true for machine learning approaches.[25] It

is particularly relevant for black box algorithms when used clinically, as there is low interpretability for clinicians in terms of how the algorithm arrives at its final prognosis. Moreover, registry data is particularly prone to reporting inaccuracies and missing data, resulting in poor prognostic ability if machine learning approaches are applied without sufficient data preparation.[26] Our approach was to combine both expert clinician manual review of the variables with automated feature selection techniques in order to arrive at the final set of variables. While time-consuming, we believe this collaborative approach is necessary in order to derive utility from registry level data. Moreover, while computationally more expensive, the ensemble machine learning approach allows for the integration of multiple types of algorithms into one cohesive model, which has been suggested to produce a more robust final product.[27]

While this is not the first study to employ machine learning techniques for OHT prediction, it describes the use of more robust feature selection techniques and the development of a larger scale ensemble ML model than has been previously reported. This example of applying modern techniques may help to overcome the registry-level data limitations that have hindered prior studies.

This study has several limitations that need to be considered when interpreting the results. First, it is retrospective in nature and subject to all inherent limitations of such studies. Most notably, there have been a number of substantial changes in the allocation system and clinical management of OHT patients over the timeframe encapsulated by the study period. As such, there is associated bias as risk models including the one developed in the current study cannot account for individual provider or transplant program decision-making. Second, the UNOS database, similar to other multicenter registries, has a number of limitations including variability in data reporting and quality. As such, assumptions are made for missing data that may introduce bias and there may be clinically important variables not captured in the available dataset. Finally, while we created a randomly selected validation cohort at the outset of the study, an independent validation cohort separate from the UNOS database was not available for testing. Further study is warranted on independent, prospective data not present in the current dataset in order to provide more comprehensive validation testing of the final model.

In conclusion, an ensemble ML model was able to achieve greater predictive performance as compared to individual ML models and logistic regression in predicting survival after OHT. This analysis demonstrates the potential of modern ML techniques in risk prediction for OHT. These approaches may have important implications in patient selection, programmatic evaluation, policy-making, and patient counseling in OHT.

Acknowledgements and Disclosures

The data reported here have been supplied by the United Network for Organ Sharing as the contractor for the Organ Procurement and Transplantation Network. The interpretation and reporting of these data are the responsibility of the authors and in no way should be seen as an official policy of or interpretation by the OPTN or the U.S. Government. Arman Kilic is on the Medical Advisory Board for Medtronic, Inc. Igor Gosev is a consultant for Abbott, Inc. Sunil Prasad is on the Scientific Advisory Board for Abbott, Inc.

References:

1. Chambers DC, Cherikh WS, Harhay MO, et al. The International Thoracic Organ Transplant Registry of the International Society for Heart and Lung Transplantation: Thirty-sixth adult lung and heart-lung transplantation Report-2019; Focus theme: Donor and recipient size match. *J Heart Lung Transplant.* 2019;38(10):1042-1055.

2. Hsich EM. Matching the Market for Heart Transplantation. *Circ Heart Fail.* 2016;9(4):e002679.

3. Jay C, Schold JD. Measuring transplant center performance: The goals are not controversial but the methods and consequences can be.*Curr Transplant Rep.* 2017;4(1):52-58.

4. Mehra MR, Uriel N, Naka Y, et al. A Fully Magnetically Levitated Left Ventricular Assist Device - Final Report. *N Engl J Med.*2019;380(17):1618-1627.

5. Fernandez FG, Shahian DM, Kormos R, et al. The Society of Thoracic Surgeons National Database 2019 Annual Report. *Ann Thorac Surg.*2019.

6. Milano CA, Rogers JG, Tatooles AJ, et al. HVAD: The ENDURANCE Supplemental Trial. *JACC Heart Fail.* 2018;6(9):792-802.

7. Weiss ES, Allen JG, Kilic A, et al. Development of a quantitative donor risk index to predict short-term mortality in orthotopic heart transplantation. *J Heart Lung Transplant.* 2012;31(3):266-273.

8. Hong KN, Iribarne A, Worku B, et al. Who is the high-risk recipient? Predicting mortality after heart transplant using pretransplant donor and recipient risk factors. *Ann Thorac Surg.* 2011;92(2):520-527; discussion 527.

9. Weiss ES, Allen JG, Arnaoutakis GJ, et al. Creation of a quantitative recipient risk index for mortality prediction after cardiac transplantation (IMPACT). *Ann Thorac Surg.* 2011;92(3):914-921; discussion 921-912.

10. 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(3):e0118644.

11. Obermeyer Z, Emanuel EJ. Predicting the Future — Big Data, Machine Learning, and Clinical Medicine. *N Engl J Med.*2016;375(13):1216-1219.

12. Cai J, Luo J, Wang S, Yang S. Feature selection in machine learning: A new perspective. *Neurocomputing.* 2018;300:70-79.

13. Gulli A, Pal S. *Deep Learning with Keras.* Packt Publishing; 2017.

14. Geissler HJ, Hölzl P, Marohl S, et al. Risk stratification in heart surgery: comparison of six score systems. *Eur J Cardiothorac Surg.* 2000;17(4):400-406.

15. O'Brien SM, Feng L, He X, et al. The Society of Thoracic Surgeons 2018 Adult Cardiac Surgery Risk Models: Part 2-Statistical Methods and Results. *Ann Thorac Surg.* 2018;105(5):1419-1428.

16. Kilic A, Weiss ES, Allen JG, et al. Simple score to assess the risk of rejection after orthotopic heart transplantation. *Circulation.*2012;125(24):3013-3021.

17. Cucchetti A, Vivarelli M, Heaton ND, et al. Artificial neural network is superior to MELD in predicting mortality of patients with end-stage liver disease. *Gut.* 2007;56(2):253-258.

18. Ayers B, Wood K, Gosev I, Prasad S. Predicting Survival after Extracorporeal Membrane Oxygenation using Machine Learning. *Ann Thorac Surg.* 2020.

19. Kilic A. Artificial Intelligence and Machine Learning in Cardiovascular Health Care. *Ann Thorac Surg.*2020;109(5):1323-1329.

20. Kilic A, Goyal A, Miller JK, et al. Predictive Utility of a Machine Learning Algorithm in Estimating Mortality Risk in Cardiac Surgery.*Ann Thorac Surg.* 2020;109(6):1811-1819.

21. Kilic A, Goyal A, Miller JK, Gleason TG, Dubrawksi A. Performance of a Machine Learning Algorithm in Predicting Outcomes of Aortic Valve Replacement. *Ann Thorac Surg.* 2020.

22. Aleksova N, Alba AC, Molinero VM, et al. Risk prediction models for survival after heart transplantation: A systematic review. *Am J Transplant.* 2020;20(4):1137-1151.

23. Medved D, Ohlsson M, Höglund P, Andersson B, Nugues P, Nilsson J. Improving prediction of heart transplantation outcome using deep learning techniques. *Sci Rep.* 2018;8(1):3613.

24. Yoon J, Zame WR, Banerjee A, Cadeiras M, Alaa AM, van der Schaar M. Personalized survival predictions via Trees of Predictors: An application to cardiac transplantation. *PLoS ONE.*2018;13(3):e0194985.

25. Adibuzzaman M, DeLaurentis P, Hill J, Benneyworth BD. Big data in healthcare - the promises, challenges and opportunities from a research perspective: A case study with a model database. *AMIA Annu Symp Proc.* 2017;2017:384-392.

26. Miller PE, Pawar S, Vaccaro B, et al. Predictive Abilities of Machine Learning Techniques May Be Limited by Dataset Characteristics: Insights From the UNOS Database. *J Card Fail.* 2019;25(6):479-483.

27. Granitto PM, Verdes PF, Ceccatto HA. Neural network ensembles: evaluation of aggregation algorithms. *Artificial Intelligence.*2005;163(2):139-162.

Figure Legends:

Figure 1. Study cohort.

Figure 2. Schematic of final ensemble ML model training and internal validation methodology.

Figure 3. Area under receiver-operating-characteristic curve comparison of each model's prognostic ability.

Figure 4. Decision curve analysis demonstrating improved risk prediction of the full ensemble ML model across the entire spectrum of predicted risk scores.

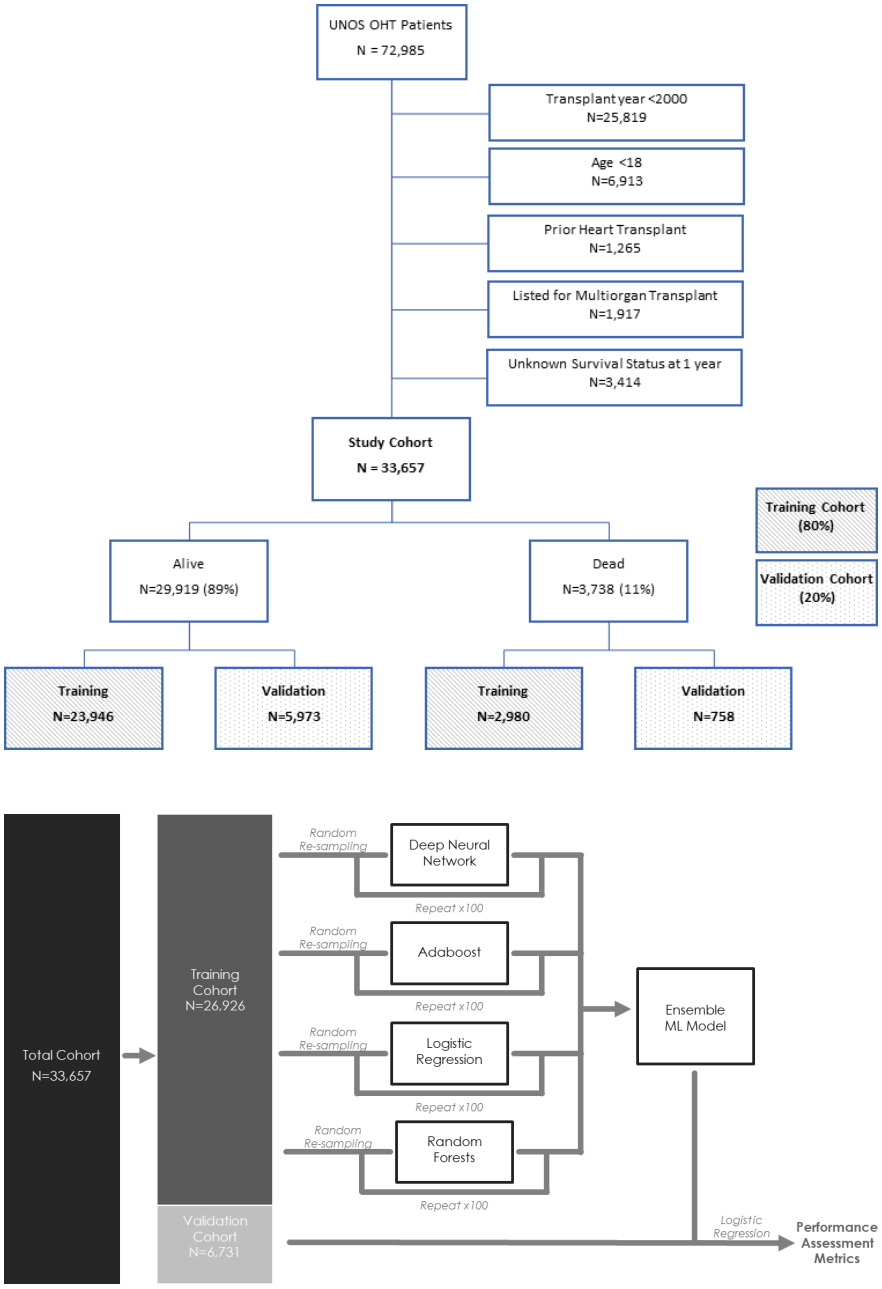
Table 1. Patient Characteristics of Training and Testing Cohorts

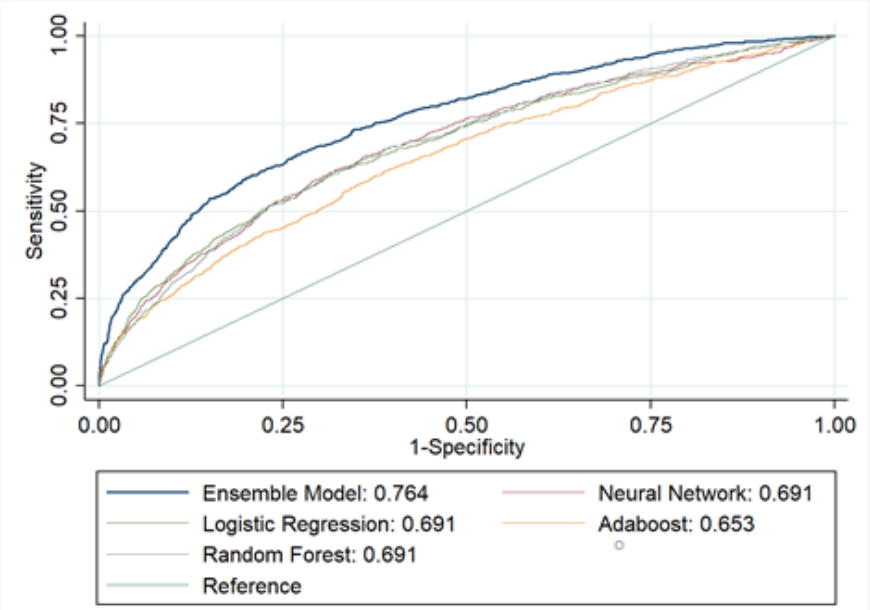
		Training (n=26,926)	**% or SD**	**Validation (n=6,731)**	**Validation (n=6,731)**	**% or SD**	**p-value**
Recipient	**Recipient**						
	Age (years)	52.8	12.5	12.5	52.8	12.3	0.873
	Female	6683	25%	25%	1706	25%	0.378
	Height (cm)	173.9	9.8	9.8	173.7	9.9	0.099
	BMI (kg/m^2)	27.0	4.8	4.8	27.1	4.8	0.453
	White	18819	70%	70%	4723	70%	0.666
	Cardiomyopathy						
	Non-ischemic	12815	48%	48%	3239	48%	0.445
	Ischemic	10986	41%	41%	2703	40%	0.339
	Congenital	715	3%	3%	196	3%	0.257
	Restrictive	670	2%	2%	171	3%	0.793
	Valvular	509	2%	2%	130	2%	0.842
	Hypertrophic	580	2%	2%	152	2%	0.607
	Prior Cardiac Surgery	12012	45%	45%	2975	44%	0.546
	History of Dialysis	596	2%	2%	169	3%	0.144
	Karnofsky Functional Status >70%	6748	25%	25%	1704	25%	0.671
	Support: Ventilator	506	2%	2%	115	2%	0.389
	Support: ICU	7789	29%	29%	1967	29%	0.631

		Training (n=26,926)	% or SD	Validation (n=6,731)	Validation (n=6,731)	% or SD	p-value
	Support: Inotropes	11037	41%	41%	2694	40%	0.149
	Support: ECMO	127	0%	0%	38	1%	0.329
	Support: IABP	1576	6%	6%	379	6%	0.503
	Mechanical Circulatory Support						
	LVAD	7465	28%	28%	1908	28%	0.309
	RVAD	50	0%	0%	9	0%	0.419
	TAH	208	1%	1%	43	1%	0.268
	LVAD+RVAD	599	2%	2%	149	2%	1.000
	Total Bilirubin	1.1	2.0	2.0	1.1	1.6	0.271
	Creatinine	1.3	0.6	0.6	1.3	0.6	0.713
	Total days on waiting list	224.0	369.9	369.9	225.1	369.2	0.836
Donor	**Donor**						
	Age (years)	31.8	11.9	11.9	31.9	12.1	0.413
	Female	7830	29%	29%	1986	30%	0.491
	Height (cm)	174.3	9.6	9.6	174.1	9.7	0.290
	BMI (kg/m^2)	26.9	5.6	5.6	27.0	5.8	0.336
	Mechanism of Death						
	Trauma	14513	54%	54%	3634	54%	0.902
	Cerebrovascular	6859	25%	25%	1741	26%	0.512
	Drug Overdose	1892	7%	7%	456	7%	0.487
	Other	3662	14%	14%	900	13%	0.633
	Blood Infection	2081	8%	8%	486	7%	0.166
Other	**Other**						
	Sex matched	19891	66%	66%	4987	74%	0.721
	Ischemic time (hours)	3.2	1.0	1.0	3.2	1.0	0.493
	Blood Type Incompatible	3931	15%	15%	966	14%	0.615

		Training (n=26,926)	% or SD	Validation (n=6,731)	Validation (n=6,731)	% or SD	p-value
	Transplant year 2000-2010	13149	49%	49%	3365	50%	0.091

BMI: body mass index; IABP: intra-aortic balloon pump; ICU: intensive care unit; ECMO: extracorporeal membrane oxygenation; LVAD: left ventricular assist device; MCS: mechanical circulatory support; RVAD: right ventricular assist device; TAH: total artificial heart





Model Type	C-Statistic	95% Confidence Interval	p-value
Full Ensemble Model	0.764	(0.745, 0.782)	*reference*
Neural Network Ensemble	0.691	(0.671, 0.712)	<0.0001
Logistic Regression Ensemble	0.691	(0.670, 0.712)	<0.0001
Adaboost Ensemble	0.653	(0.632, 0.674)	<0.0001
Random Forest Ensemble	0.691	(0.671, 0.711)	<0.0001
Logistic Regression (singular)	0.649	(0.628, 0.670)	<0.0001