IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 26 March 2024, accepted 5 June 2024, date of publication 24 June 2024, date of current version 5 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3418146

RESEARCH ARTICLE

Improving 1-Year Mortality Prediction After Pediatric Heart Transplantation Using Hypothetical Donor-Recipient Matches

MICHELA VENTURINI[1,2,*], FATEME NATEGHI HAREDASHT[1,2,3,*], FRANTIŠEK SABOVČIK[4], ROBERT J. H. MILLER[5], TATIANA KUZNETSOVA[4], AND CELINE VENS[1,2]

[1]KU Leuven, Campus KULAK, Department of Public Health and Primary Care, Etienne Sabbelaan 53, 8500 Kortrijk, Belgium
[2]ITEC, imec and KU Leuven, Etienne Sabbelaan 51, 8500 Kortrijk, Belgium
[3]Stanford Center for Biomedical Informatics Research, Stanford University, CA, USA
[4]KU Leuven, Department of Cardiovascular Sciences, Herestraat 49, 3000 Leuven, Belgium
[5]Division of Cardiac Sciences, Libin Cardiovascular Institute of Alberta, University of Calgary, Calgary, Canada

Corresponding author: Michela Venturini (michela.venturini@kuleuven.be)

This work was supported in part by the Health Resources and Services Administration under Contract HHSH250-2019-00001C, in part by the Research Fund Flanders under Project G0A2120N, and in part by Katholieke Universiteit (KU) Leuven Internal Funds under Grant 3M180314.

*Michela Venturini and Fateme Nateghi Haredasht contributed equally to this work.

ABSTRACT Heart transplantation is a life-saving procedure for children affected by end-stage heart failure. However, despite recent improvements in long-term outcomes, 1-year post-transplantation mortality has remained relatively high. Accurate prediction of post-transplantation mortality is crucial to evaluating risks related to recipient-donor matches. Machine learning techniques can potentially improve the current allocation system through the integration of a larger set of features. In this work, we improve 1-year mortality prediction after pediatric heart transplantation using a new self-training approach, based on generating artificial recipient-donor pairs as synthetic unlabeled observations. We tested and compared our approach to several baselines using the cohort of pediatric patients in the UNOS database. Our study suggests that augmenting the dataset with proper synthetic observations can improve the prediction of 1-year mortality after pediatric heart transplantation. Our findings have implications for the future of heart transplantation in children, offering a potential path to refine recipient-donor matching and improve survival rates. This study contributes to the growing field of advanced machine learning techniques applied to medical decision-making, specifically in the context of organ transplantation.

INDEX TERMS Heart transplantation, machine-learning, post-transplantation mortality, self-training, synthetic observations.

I. INTRODUCTION

Heart transplantation (HTx) has become a life-saving procedure for children affected by end-stage heart failure. Although pediatric HTx only represents roughly 10% of the total number of HTx, their number has consistently grown during the past decades with more than 450 transplantations in 2020 in the U.S.. Despite the noticeable improvement in long-term outcomes after pediatric HTx, 1-year post-transplantation recipient mortality has remained relatively high [7]. In addition, waitlist mortality for pediatric HTx still represents a crucial issue, especially due to the limited organ supply, since many pediatric heart donors are discarded. Efforts have been made to support physicians to efficiently determine acceptable heart quality and ultimately increase donor organ utilization, especially by investigating factors related to the transplant outcome, as well as by improving data visualization tools [4], [10]. It is worth

The associate editor coordinating the review of this manuscript and approving it for publication was Mohammad Zia Ur Rahman.

mentioning that the decision-making process remains highly variable and dependent on several variables. Thus, accurate prediction of the post-transplantation outcome is central to optimizing the organ allocation system and allowing for more informed medical decision-making [19].

Prediction models for heart transplantation allocation are already in use. One of the most commonly used models is the Heart Transplantation Survival Score (HTSS), which was developed by the United Network for Organ Sharing (UNOS) in the United States [13]. The HTSS takes into account several factors, including the recipient's age, diagnosis, functional status, and the presence of other medical conditions, such as diabetes or kidney disease. The model also considers the donor's age, cause of death, and whether the donor and recipient have compatible blood types. Based on these factors, the HTSS assigns each potential recipient a score that reflects their predicted survival after a heart transplant. Transplant centers use this score to help determine the priority of patients on the waiting list. Another example of a model developed for heart transplant allocation is the Eurotransplant International Foundation's Eurotransplant Donor Risk Index (ET-DRI) [2]. Similarly to the previous one, this model considers several donor and recipient factors to help guide transplant allocation decisions.

Despite the potential usefulness of machine learning (ML) in this context, most of the currently adopted strategies are based on simpler strategies. The HTSS, for example, is based on a statistical analysis of data collected from thousands of heart transplant recipients in the United States. The model uses a set of predetermined variables and weights to calculate a recipient's predicted survival after a transplant. Similarly, the ET-DRI is based on a statistical analysis of data from heart transplant recipients in Europe. The outcome of pediatric HTx is affected by a large number of variables, related both to the donor and recipient [8], making ML models a valid alternative to multivariate regression techniques for HTx outcome prediction, as they can handle a higher number of variables as well as complex non-linear relationships between them. ML models have already been applied, to both adult and pediatric cohorts, and the recent literature suggests that ML models are generally more accurate in predicting mortality compared to standard regression techniques [15], [17]. However, the majority of applied retrospective validation techniques do not take into account the time shift between observations and population or policy changes, which likely results in an overestimation of the performance, as shown by [16]. Furthermore, the relatively small sample size of pediatric cohorts compared to the adult cohort, quality of data, and inter-center variability represent a limitation to predictive performance and model generalizability [9].

In this paper, we aim to leverage techniques from the semi-supervised learning domain to HTx outcome prediction and validate the performance considering the time shift between observations. We hypothesize that predictive performance can be improved by enlarging the training set with hypothetical donor-recipient matches. Thus, we re-combine donors and recipients from the training set and use them as unlabeled observations. Semi-supervised learning (SSL) is a type of machine learning where a model is trained using a combination of labeled and unlabeled data. SSL can be particularly beneficial to improve a model in case of scarce labeled observations or when unlabeled observations contain additional relevant information [18]. SSL has already been applied in medicine in the context of diagnosis prediction with promising results under precise assumptions [6]. Self-training is a well-known SSL technique that uses a first model trained exclusively on labeled data to iteratively predict labels of unlabeled observations and add the most confident predictions to the training set, to improve predictive capabilities. To the best of our knowledge, this technique has never been applied in the setting of post-transplantation outcome prediction. In our work, we exploit the artificially generated donor-recipient matches as unlabeled observations in a self-training framework to enhance 1-year post-transplantation mortality prediction in a pediatric cohort. We evaluate our work on a real dataset of around ten thousand patients.

II. MATERIALS AND METHODS

A. DATASET

We included pediatric patients (age $<$ 18 years old) enrolled in the United Network of Organ Sharing (UNOS) database [5] between January 1994 and December 2016 who underwent cardiac transplantation. Donor, recipient, and matching variables included in the ML models were measured at the time of listing or at the time of transplantation. This study complied with the Declaration of Helsinki and the ISHLT ethics statement and utilized publicly available data.

B. OUR FRAMEWORK

The original dataset consists of individual heart transplantation events, with features falling into three categories: donor features, recipient features, and donor-recipient match features (such as weight ratio). The structure of the dataset itself allows to build synthetic instances (or matches) by merging donor features with different recipient features and calculating the new match features. The outcome of interest is the 1-year mortality for the recipient, and we assume that it depends on all three feature categories. In our work, we developed a two-step framework, which (1) generated a set of plausible synthetic donor-recipient matches and (2) employed self-training, depicted in Fig. 1, to improve the predictive performance of the model. We evaluated the framework's performance by taking into account the temporal shift in the dataset, by using rolling cross-validation and therefore providing a more realistic evaluation.

1) OBTAINING SYNTHETIC OBSERVATIONS

An important assumption of SSL is the smoothness assumption (if two samples x and x' are close in the input space, their labels y and y' should be the same) [18]. Including such an assumption may reduce the possibility of generating clinically unreliable matches. Therefore, we populated

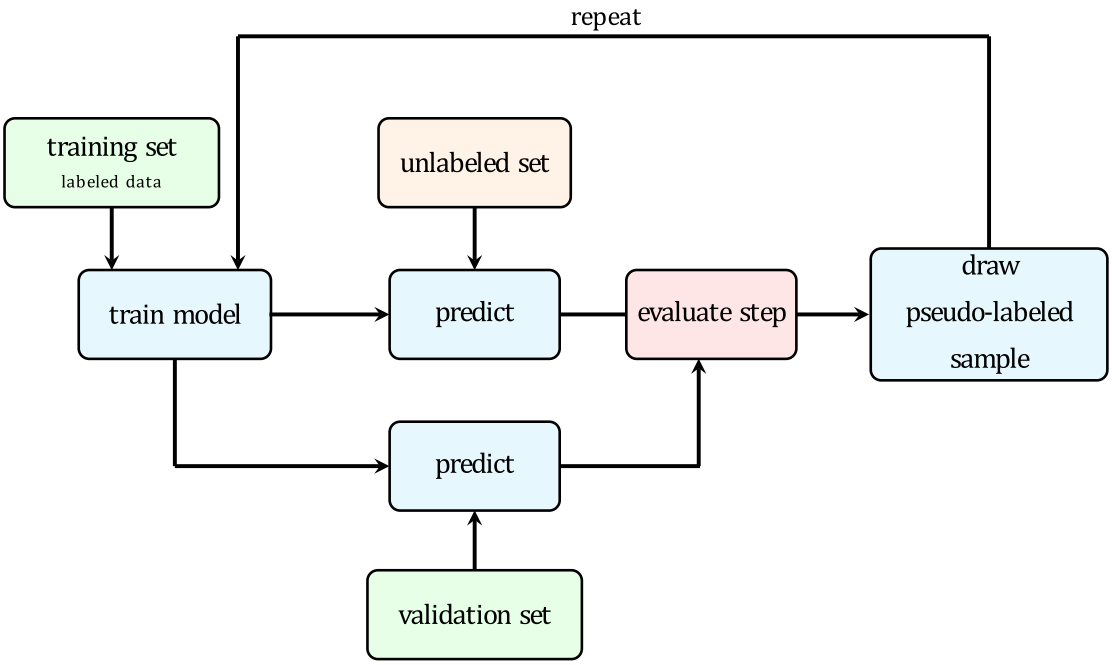

FIGURE 1. Representation of the self-training step of the framework.

our unlabeled set of synthetic observations by creating donor-recipient matches as similar as possible to the original ones. Firstly, we used the K-Nearest Neighbor (K-NN) algorithm to find, for each existing match, the nearest neighbors (matches) in the dataset. Consequently, we built the new matches by concatenating each donor with the recipients of the nearest matches. We also calculated the matching features relative to the new matches, such as height and weight ratio, as they represent a valuable source of information to determine the quality of a match (see Fig. 2).

2) FILTERING STEP

Additionally, we performed a "filtering" step to select a set of more reliable synthetic observations. Specifically, we trained a survival model on the labeled set, considering time-to-death as the outcome, and we predicted the mortality for the unlabeled set. Survival models consider not just whether an event occurs, but also when it occurs. This approach leverages the temporal dimension, providing a richer and more informative analysis of the event's occurrence over time. The predicted mortality was then used to select two types of artificial sets. In one setting, we used such value to select those synthetic observations with the lowest predicted chance of dying (observations whose mortality fell within the 25^{th} percentile of all the predictions), resulting in a set of "negative" synthetic observations, in the sense that we could assume they represented a population that would not experience death. In the other setting, we performed the same procedure to select the synthetic observations with the highest chances of dying, resulting in a set of "positive" synthetic observations.

3) SELF-TRAINING FRAMEWORK

In contrast to the conventional self-training approach, where the unlabeled observations may belong to either class, our framework assumes that the synthetic observations are categorized as either positive (i.e., patients who died within 1-year post-transplantation) or negative, and leverages self-training to incorporate the most confident predictions. In the initial iteration of our framework, we trained a binary classifier exclusively on labeled observations. Subsequently, the model predicted the probability of death for randomly selected batches of synthetic negative or positive observations in each iteration. The most reliable observations were then incrementally incorporated into the training set. The number of iterations was determined based on the performance on a validation set, and the optimal iteration was selected. It is worth mentioning that the synthetic observations were not used to evaluate the model's performance.

4) BASE LEARNERS

Our methodology relied on tree-based models, specifically Random Survival Forest (RSF) [11] for the filtering step and Random Forest Classifier (RF) [3] for the self-training. RSF considers time-to-event information as the outcome, and the splitting rule is based on the log-rank test which evaluates how different are the estimated survival curves in the resulting nodes. RF accepts a binary outcome and the splitting rule is based on the Gini index. Tree-based models are capable of handling complex non-linear relationships between variables and provide state-of-the-art performance on tabular data. Moreover, they have been shown to provide superior or competitive predictive performance for post-transplantation mortality, compared to alternative techniques [16].

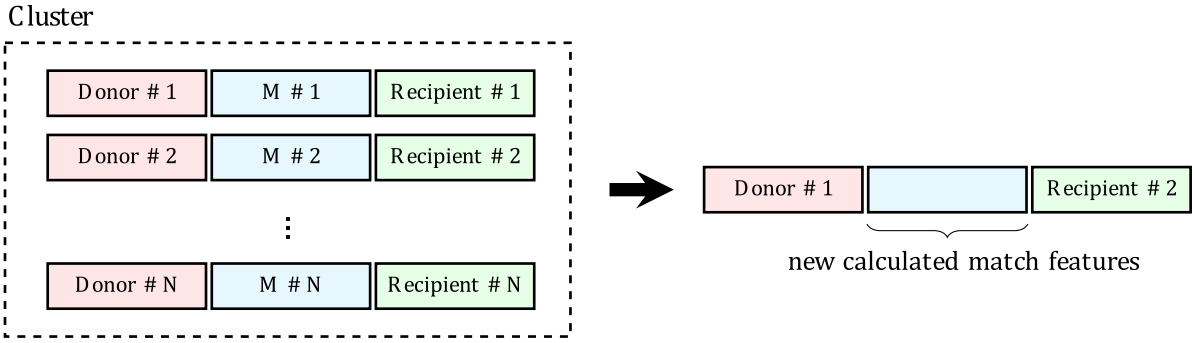

FIGURE 2. Illustration of the process generating the unlabeled set using K-NN. The left figure shows the structure of the data (cluster) where M stands for the match features between the donor-recipient pair. Using K-NN, we find the best new donor-recipient pair after which we calculate the new match features.

5) EVALUATION CRITERION

To select the most reliable observations and determine the best iteration of our framework, we defined the performance gain G at iteration i (with $i = 1, 2, \ldots, N$) as follows:

$$G_i = \text{AUPR}_i - \text{AUPR}_{i-1}, \tag{1}$$

with AUPR being the area under the precision-recall curve. G_i was used at each iteration to decide whether to add or discard the batch of synthetic observations, based on a threshold α, indicating the minimum accepted gain. The best iteration was selected based on the cumulative gain cG_i between iteration 0 and iteration i. The choice of AUPR to define the gain was dictated by the unbalanced nature of the dataset towards the observations free from the event, as AUPR is a more appropriate measure than other evaluation metrics as it is better suited to evaluate the model's ability to identify observations in the minority class.

6) ROLLING CROSS-VALIDATION

For evaluating the performance of our proposed framework, we employed the temporal rolling cross-validation method similar to [16], which is a technique specifically employed when the data exhibit temporal or time-based characteristics. In essence, this technique entails dividing the data into multiple folds, where each fold comprises a contiguous subset of the data. The model is trained on the first fold and evaluated on the second fold, then trained on the first and second folds and evaluated on the third fold, and so on, until all folds have been utilized as the testing set once. Specifically in our study, this approach involves treating each year of transplantation as a distinct fold. For each fold, the training set comprises past transplant data that occurred before the current year under evaluation, whereas the test set represents future data containing patients transplanted during the current year. As a result, the models are trained on data from a different time period than the test data. A validation set of one year after the training set has also been used to select the best number of iterations in the fold. The final outcome is determined by combining the predictions for each test year and computing the averaged AUROC and AUPR. By following this rolling process, temporal rolling cross-validation guarantees that the model is assessed on data that resembles the data it will encounter in the future, thus providing a more realistic estimate of the model's predictive ability. The stability of model performance is another valuable aspect of temporal rolling cross-validation. It provides multiple evaluation metrics across different periods, yielding a more robust estimate of the model's performance and helping to identify any variations in predictive ability over time. Another important advantage of this approach is the opportunity for data-driven model updates. As new data becomes available, the model can be retrained periodically to incorporate the latest information, ensuring its continued relevance and accuracy as the dataset expands. Furthermore, this technique precludes the risk of data leakage from the future into the past, which can lead to overly optimistic performance estimates (Fig. 3).

III. EXPERIMENTAL SETUP

For the first iteration of the rolling cross-validation, we employed the observations belonging to the first 10 years of the dataset as the training set. At each iteration, we obtained the synthetic set of observations by exclusively using the training set. Subsequently, we run K-NN with Euclidean distance and we set the number of neighbors to four. We removed the censored observations (with a follow-up lower than 365 days) from the train, validation, and test set to train the RF classifier and evaluate the predictive performance. We evaluated the AUPR at each iteration by bootstrapping 1000 samples from the validation set. We used the bootstrapped estimate of the performance to calculate the gain, as defined in Equation 1. We chose a negative $\alpha = -0.005$ to avoid over-fitting on the validation set, thus allowing for a small decrease in performance when adding the batch of synthetic observations. We selected the best iteration based on the highest cG_i on the validation set. We compared our framework (SSL_OUR) with several baselines, the simplest being two models, Linear regression and RF, trained without unlabeled observations (LR_NO_U and RF_NO_U), and an RF trained on labeled and all the synthetic observations added at once (RF_ALL_U). We considered two different settings for our framework, the first using the set of negative synthetic observations (referred

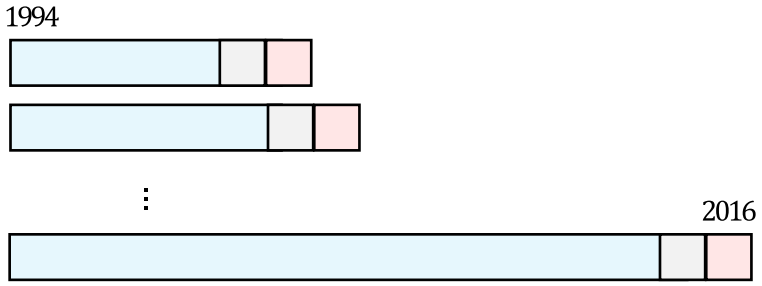

FIGURE 3. To apply the rolling cross-validation approach, patients were added incrementally to the training set according to their year of transplantation. The blue block in the figure illustrates the initial training data. Subsequently, the data from the following year was used as a validation set, represented by the grey block in the figure. The test set, depicted by the red block in the figure, was collected two years after the training set.

TABLE 1. Cohort characteristics by the outcome. The outcome was defined as the patient's status at one year after HTx. All values are reported in the median and interquartile ranges unless specified. ECMO: extracorporeal membrane oxygenation. *P-value < 0.001.

	Alive N=7838	Dead N=961
Recipient characteristics		
Age*	7.0 [1.0,14.0]	2 [0.0,12.0]
Female, n (%)	3445 (44.0)	444 (46.2)
Retransplant, n (%)	531 (6.8)	90 (9.4)
ECMO*, n (%)	312 (4.0)	142 (14.8)
Creatinine at listing (mg/dL)*	0.5 [0.4,0.8]	0.6 [0.4,0.8]
Bilirubin (mg/dL)*	0.7 [0.4,1.2]	0.9 [0.5,2.1]
Donor characteristics		
Age*	9.0 [1.0,16.0]	4.0 [0.0,14.0]
Female, n (%)	3236 (41.3)	423 (44.0)
Insuline dependency, n (%)	45 (0.6)	2 (0.2)
History of malignancy*, n (%)	159 (2.0)	14 (1.5)
Ischemic time (minutes)*	209.0 [165.0,250.0]	221.0 [166.0,266.0]

to as (N)) and the second using the positive one (referred to as (P)). All experiments were performed in Python, version 3.8. All models were fitted using the scikit-learn[1] and scikit-survival[2] Python packages with default parameters.

IV. RESULTS AND DISCUSSION

A. COHORT CHARACTERISTICS

The original pediatric cohort included 9236 pediatric patients. We excluded patients undergoing combined heart and lung transplants (n=201), patients with missing information about the outcome (n=146), and patients with less than 80 available variables (n=90). This led to a final cohort of 8799 patients, with approximately 11% of them dying within a year post-HTx. The characteristics of our cohort are presented in Table 1. The Kruskal-Wallis and Pearson's chi-squared test was used for non-parametric comparisons between continuous and discrete variables, respectively.

B. PREDICTIVE PERFORMANCE

The outcomes of our methodology compared to the considered baselines are reported in Table 2. Our SSL_OUR (N) framework exhibited better performance than the baselines concerning AUROC and AUPR, with the highest improvement observed in the AUROC metric. Additionally, our approach outperformed the method of adding all synthetic observations simultaneously, as our self-training process inherently selects the most dependable observations to be incorporated into the training set, discarding noisy ones. The performance reported for our methodology surpasses that of all models documented in [16], where the authors utilized a similar pre-processing approach and the same rolling cross-validation technique. Fig. 4 illustrates the key statistics obtained from the rolling cross-validation process. The top-left panel shows the optimal iteration of the SSL framework. In contrast, the top-right panel displays the size of the training set at the best iteration compared to the size of the training set without synthetic observations. The bottom panel presents the AUROC and AUPR metrics computed on the test set for the best iteration model and the model trained without synthetic observations. The decline in performance, particularly towards the final iteration, can be attributed to two primary factors. Firstly, the number of 1-year post-transplant deaths recorded in the dataset gradually decreases over the years, from 70 (31.4%) in 1994 to only 34 in 2016 (4.6%). Secondly, several changes to the pediatric heart allocation policy were introduced in 2016.[3] Fig. 5 reports the execution of our framework taken from two different steps of the rolling cross-validation. The top figures regard step #2, whereas the bottom ones regard step #9. In the first scenario, the best iteration is the 27^{th}, when roughly 500 synthetic observations were added to the training set. The gain in AUPR for the test set is roughly 6% and the best gain for the validation set corresponds to the best gain for the test

TABLE 2. Performance evaluated by rolling cross-validation. Reported values indicate the mean, with the best values in bold.

Method	AUROC	AUPR
LR_NO_U	0.677	0.181
RF_NO_U	0.728	0.299
RF_ALL_U (P)	0.664	0.183
RF_ALL_U (N)	0.723	0.255
SSL_OUR (P)	0.724	0.283
SSL_OUR (N)	**0.745**	**0.309**

[1]https://scikit-learn.org/stable/ (Accessed: 11/04/2023)
[2]https://scikit-survival.readthedocs.io/en/stable/ (Accessed: 11/04/2023)
[3]https://optn.transplant.hrsa.gov/professionals/by-organ/heart-lung/pediatric-heart-allocation/ (Accessed: 11/04/2023)

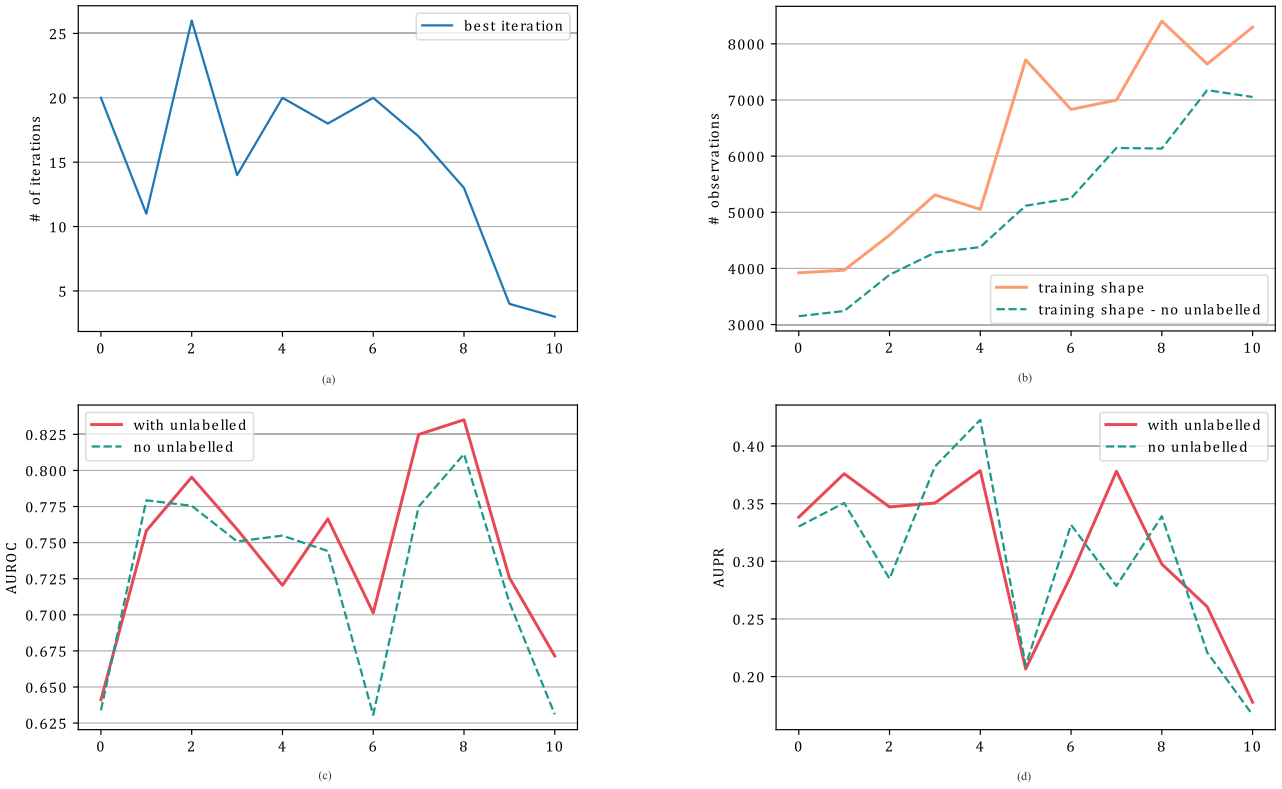

FIGURE 4. From the top-left. Best iteration, training shape, test AUROC, and test AUPR, for each of the rolling cross-validation steps.

set. On the other hand, in the second scenario, the best gain is reached at the 5^{th} iteration, with roughly 400 synthetic observations added. In this case, the gain in AUPR on the test set is more modest. However, in both scenarios, we observe that the best iteration according to the AUPR calculated on the validation set, corresponds to the best iteration also on the test set, indicating that the framework is effective in adding reliable synthetic observations.

C. FEATURE IMPORTANCE

We extracted the 20 most important features of Logistic Regression and Random Forest without unlabelled observations, and compared them with the outcome of our proposed model, by using SHAP values [14]. The results of this comparison are illustrated in Fig. 6. Important features identified by all models include both donor and recipient features (e.g., height, weight, and age). On the other hand, matching features were identified as important only by RF models. Features identified as important exclusively by our model are sex match and donor creatinine. Additionally, we report the SHAP summary plot (Fig. 7) of our model, to better visualize the impact of the variables on the model's predictions.

D. PRIOR RESEARCH IN PEDIATRIC HEART TRANSPLANT PREDICTION

It is essential to contextualize our findings within the broader landscape of machine learning applications in pediatric heart transplantation prediction. Our study aligns well with the recent work by [12], in which they ventured into machine learning to forecast what unfolds after pediatric heart transplantation. Their focus on innovative approaches to improve pediatric care resonates with our own efforts. Additionally, [1] took a deep dive into predicting one-year mortality using machine learning, emphasizing factors such as total serum bilirubin, BMI, and Sgpt. Our findings complement theirs, shedding light on the multifaceted factors that influence post-transplantation outcomes. Moreover, renal insufficiency markers (such as creatinine), recipient BMI, recipient total bilirubin, and diagnosis before transplantation are known as important factors in determining patients' survival after pediatric heart transplant [8]. Finally, our approach is consistent with the exploration carried out by [15], which investigated the utility of machine learning in predicting mortality after pediatric heart transplantation. While they encountered some sensitivity challenges, our methodology takes a step forward in enhancing predictive accuracy. Overall, these studies illustrate the growing interest in employing machine learning for risk assessment and outcome prediction in pediatric heart transplantation. This trend holds promise for refining patient stratification and enhancing post-transplant care within the field.

V. LIMITATIONS

In our study, we exclusively focus on the pediatric cohort from the UNOS database, and an interesting direction would be to repeat the procedure on the adult cohort to further validate the results. The potential differences between

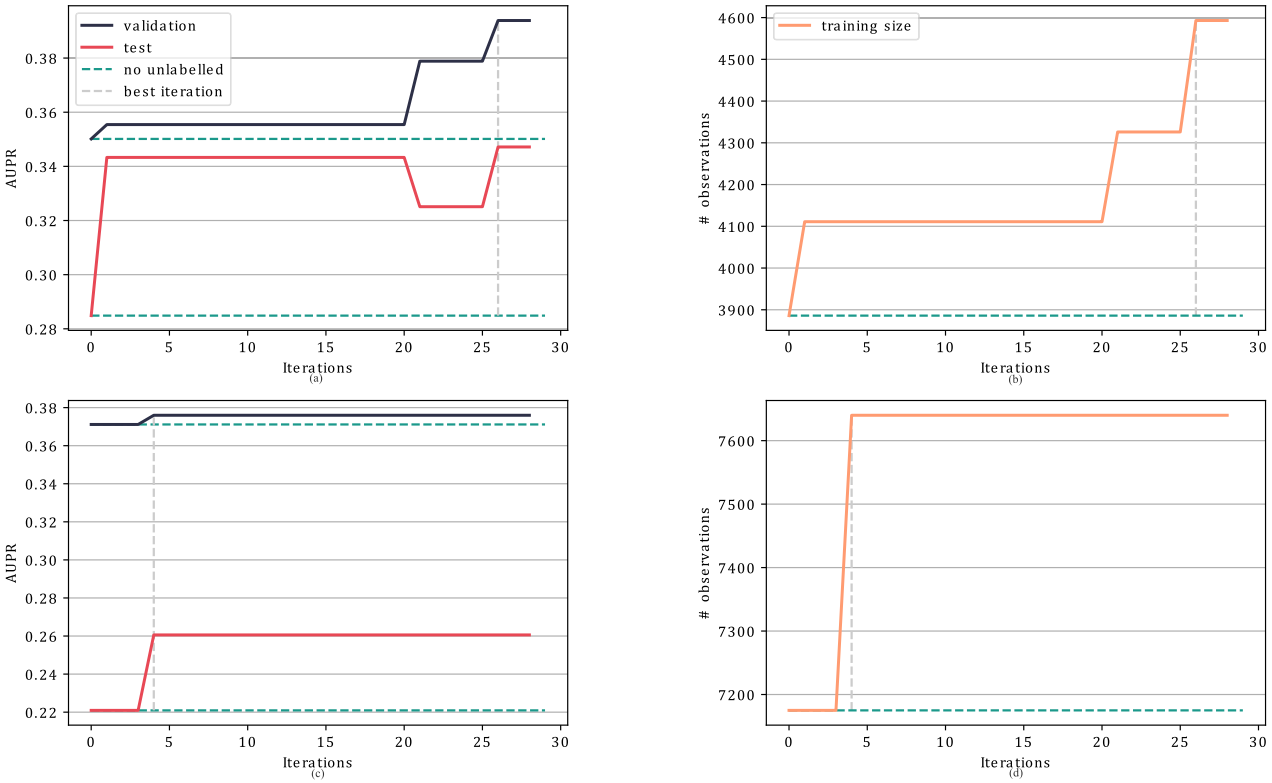

FIGURE 5. Performance and training set size across self-training iterations. Two cases are depicted: the top figures regard step #2 of the rolling cross-validation, whereas the bottom figures regard step #9.

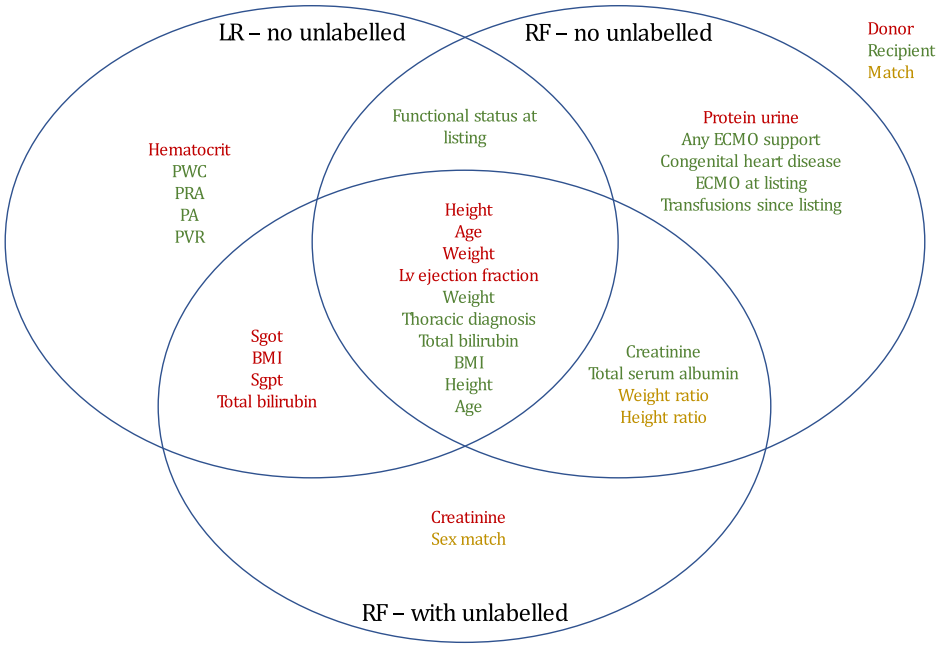

FIGURE 6. Models' most important features. PWC Physical Work Capacity, PRA Panel Reactive Antibody calculated, PA Pulmonary Artery diastolic pressure, PVR Pulmonary Vascular Resistance, BMI Body Mass Index, Sgot Serum Glutamic-Oxaloacetic Transaminase, Sgpt Serum Glutamic Pyruvic Transaminase.

pediatric and adult populations could offer valuable insights into the generalizability of our approach across different age groups.

Moreover, we exclusively focused on tree-based models, as the previous literature supported their state-of-the-art performance in this application. While tree-based models have shown promising results, future research could explore the performance of other machine learning algorithms to compare their predictive capabilities for pediatric heart transplantation outcomes.

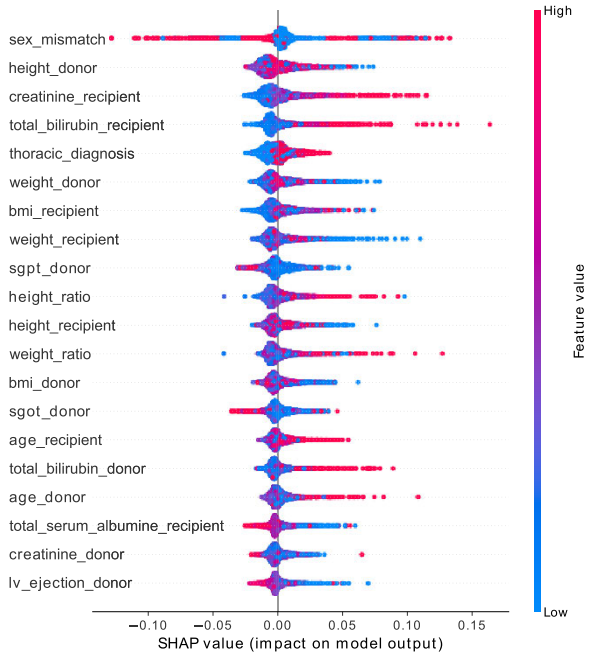

FIGURE 7. SHAP value of Random Forest model output. Each point represents a variable together with an observation. As demonstrated by the color bar, higher values are shown in red, whereas lower values are shown in blue.

Furthermore, it is essential to acknowledge that the dataset used in this study is from the United Network of Organ Sharing (UNOS) database and covers patients enrolled between January 1994 and December 2016. As of now, this is the most recent UNOS data that we have access to. Unfortunately, we do not have access to more recent data beyond 2016. The availability of more recent data could have provided additional insights into the evolving landscape of pediatric heart transplantation outcomes.

Additionally, we did not perform extensive parameter tuning for the models but rather referred to the default settings from the scikit-learn library. While the default settings are commonly used as a starting point, fine-tuning the model's hyperparameters could potentially further optimize the predictive performance.

Another limitation of our study is that we have not specifically investigated the model's performance for longer-term survival, such as 5-year survival, in the context of pediatric heart transplantation. Whereas predicting longer-term survival outcomes is of great importance in the field of heart transplantation, it introduces complexities due to potential changes in patient characteristics, treatment protocols over time, and the availability of long-term follow-up data. Therefore, our study focused primarily on improving the prediction of 1-year mortality, which is a critical period for post-transplant monitoring and interventions. We hope that by incorporating more extended and updated datasets, as well as advanced survival analysis techniques, our study will help to improve the understanding of post-transplant outcomes in pediatric heart transplantation through future investigations.

VI. CONCLUSION

We proposed a novel semi-supervised learning approach to enhance the accuracy of 1-year mortality prediction following pediatric heart transplantation. To ensure clinical reliability, we generated synthetic instances by pairing donors and recipients that closely resembled real-world cases. Our study demonstrates that incorporating appropriate unlabeled data into a self-training framework improves prediction performance. In future research, we plan to explore various parameters, such as the gain definition (α) and the base learner type for self-training, to further optimize our approach. Furthermore, we aim to investigate alternative clustering techniques to generate the synthetic set of observations, as this is a critical step in improving the performance of semi-supervised learning techniques.

ACKNOWLEDGMENT

The content is the responsibility of the authors alone and it does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. The authors also acknowledge the Flemish Government (AI Research Program).

REFERENCES

[1] A. Ashfaq, G. M. Gray, J. Carapellucci, E. K. Amankwah, L. M. Ahumada, M. Rehman, J. A. Quintessenza, and A. Asante-Korang, "Predicting one year mortality using machine learning after pediatric heart transplantation: Analysis of the united network of organ sharing (UNOS) database," *J. Heart Lung Transplantation*, vol. 41, no. 4, p. S152, Apr. 2022.

[2] A. E. Braat, J. J. Blok, H. Putter, R. Adam, A. K. Burroughs, A. O. Rahmel, R. J. Porte, X. Rogiers, and J. Ringers, "The eurotransplant donor risk index in liver transplantation: ET-DRI," *Amer. J. Transplantation*, vol. 12, no. 10, pp. 2789–2796, Oct. 2012.

[3] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, Oct. 2001.

[4] J. Bullock, M. Grieco, Y. Liu, I. Pedersen, W. Roberson, G. Wright, P. Alonzi, M. A. McCulloch, and M. D. Porter, "Determining factors of heart quality and donor acceptance in pediatric heart transplants," in *Proc. Syst. Inf. Eng. Design Symp. (SIEDS)*, Apr. 2021, pp. 1–6.

[5] [Online]. Available: http://optn.transplant.hrsa.gov

[6] A. Chebli, A. Djebbar, and H. F. Marouani, "Semi-supervised learning for medical application: A survey," in *Proc. Int. Conf. Appl. Smart Syst. (ICASS)*, Nov. 2018, pp. 1–9.

[7] M. Colvin, J. M. Smith, Y. Ahn, M. A. Skeans, E. Messick, K. Bradbrook, K. Gauntt, A. K. Israni, J. J. Snyder, and B. L. Kasiske, "OPTN/SRTR 2020 annual data report: Heart," *Amer. J. Transplantation*, vol. 22, pp. 350–437, Mar. 2022.

[8] A. I. Dipchand, "Current state of pediatric cardiac transplantation," *ASVIDE*, vol. 5, pp. 1–116, Feb. 2018.

[9] N. Gotlieb, A. Azhie, D. Sharma, A. Spann, N.-J. Suo, J. Tran, A. Orchanian-Cheff, B. Wang, A. Goldenberg, M. Chassé, H. Cardinal, J. P. Cohen, A. Lodi, M. Dieude, and M. Bhat, "The promise of machine learning applications in solid organ transplantation," *NPJ Digit. Med.*, vol. 5, no. 1, pp. 1–13, Jul. 2022.

[10] C. Hyldahl, O. Kaczmarskyj, J. Laruffa, A. Miller, L. Snavely, A. Wan, and S. L. Riggs, "Designing a dashboard to streamline pediatric heart transplant decision making," in *Proc. Syst. Inf. Eng. Design Symp. (SIEDS)*, Apr. 2023, pp. 237–242.

[11] J. M. G. Taylor, "Random survival forests," *J. Thoracic Oncol.*, vol. 6, no. 12, pp. 1974–1975, Dec. 2011.

[12] M. O. Killian, S. Tian, A. Xing, D. Hughes, D. Gupta, X. Wang, and Z. He, "Prediction of outcomes after heart transplantation in pediatric patients using national registry data: Evaluation of machine learning approaches," *JMIR Cardio*, vol. 7, Jun. 2023, Art. no. e45352.
[13] J. K. Kirklin, D. C. Naftel, R. L. Kormos, L. W. Stevenson, F. D. Pagani, M. A. Miller, J. T. Baldwin, and J. B. Young, "The fourth INTERMACS annual report: 4,000 implants and counting," *J. Heart Lung Transplantation*, vol. 31, no. 2, pp. 117–126, Feb. 2012.
[14] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1–11.
[15] R. Miller, D. Tumin, J. Cooper, D. Hayes, and J. D. Tobias, "Prediction of mortality following pediatric heart transplant using machine learning algorithms," *Pediatric Transplantation*, vol. 23, no. 3, May 2019, Art. no. e13360.
[16] R. J. H. Miller, F. Sabovč ik, N. Cauwenberghs, C. Vens, K. K. Khush, P. A. Heidenreich, F. Haddad, and T. Kuznetsova, "Temporal shift and predictive performance of machine learning for heart transplant outcomes," *J. Heart Lung Transplantation*, vol. 41, no. 7, pp. 928–936, Jul. 2022.
[17] V. Naruka, A. Arjomandi Rad, H. Subbiah Ponniah, J. Francis, R. Vardanyan, P. Tasoudis, D. E. Magouliotis, G. L. Lazopoulos, M. Y. Salmasi, and T. Athanasiou, "Machine learning and artificial intelligence in cardiac transplantation: A systematic review," *Artif. Organs*, vol. 46, no. 9, pp. 1741–1753, Sep. 2022.
[18] X. Yang, Z. Song, I. King, and Z. Xu, "A survey on deep semi-supervised learning," *IEEE Trans. Knowl. Data Eng.*, vol. 109, no. 2, pp. 1–20, Aug. 2022.
[19] R. J. Williams, M. Lu, L. A. Sleeper, E. D. Blume, P. Esteso, F. Fynn-Thompson, C. J. Vanderpluym, S. Urbach, and K. P. Daly, "Pediatric heart transplant waiting times in the United States since the 2016 allocation policy change," *Amer. J. Transplantation*, vol. 22, no. 3, pp. 833–842, Mar. 2022.

MICHELA VENTURINI received the B.Sc. degree in bioinformatics from the University of Verona, in 2018, and the M.Sc. degree in data science from the University of Trieste, in 2020. She is currently pursuing the Ph.D. degree with ITEC, imec Research Group, KU Leuven, Belgium. Her research interests include machine learning applications for healthcare, explainable AI, and survival analysis.

FATEME NATEGHI HAREDASHT received the B.Sc. degree in electrical engineering from the University of Guilan, the M.Sc. degree in biomedical engineering (bioinformatics) from the Amirkabir University of Technology (Tehran Polytechnic), and the Ph.D. degree in biomedical sciences with emphasis on health and technology from KU Leuven, in 2023. She is currently a Postdoctoral Researcher with the Stanford Center for Biomedical Informatics Research. Her research interests include semisupervised learning, machine learning for healthcare, explainable AI, and survival analysis.

FRANTIŠEK SABOVČIK received the master's degree in mathematical methods in information technology from Brno University of Technology, Czech Republic. He is currently pursuing the Ph.D. degree with KU Leuven, Belgium, with a focus on improving early cardiovascular risk assessment using machine learning. His main background is in computer and data science, with a focus on biomedical applications. He has extensive experience in software development and worked in various tech companies. For example, he was a Senior Developer with ThreatMark, a bank fraud detection software, or in Figure, a New York-based start-up developing business management software. Later in his career, he shifted toward machine learning and data science. His main focus is on developing practical ML-based tools for early risk assessment of cardiovascular disease. These approaches include the detection of asymptomatic cardiac abnormalities or incident heart failure prediction in the general population.

ROBERT J. H. MILLER is currently a Clinical Associate Professor with the Department of Cardiac Sciences, University of Calgary, and the Medical Director of nuclear cardiology and cardiac CT program. He completed training in internal medicine and general cardiology with the University of Calgary. He went on to complete training in advanced heart failure and cardiac transplantation with Stanford University and advanced cardiac imaging with the Cedars-Sinai Medical Center. His clinical work also includes nuclear cardiology and cardiac CT and time on the advanced heart failure service and cardiac function clinic. His research interests include advanced cardiac imaging applications in patients with cardiomyopathies and epidemiology in heart failure and cardiac transplantation.

TATIANA KUZNETSOVA received the Ph.D. degree in biomedical science from KU Leuven, Belgium, in 2004. She is currently a Professor with the Department of Cardiovascular Sciences, KU Leuven, and the Head of the Research Unit Hypertension and Cardiovascular Epidemiology, Department of Cardiovascular Sciences, KU Leuven. Her research interests include integrative cardiovascular phenomics, cardiorespiratory fitness, and biomarker profiling for the assessment of cardiac health in the community.

CELINE VENS received the Ph.D. degree in computer science (machine learning) from KU Leuven, Belgium, in 2007. She is currently a Professor with the Faculty of Medicine, KU Leuven, and ITEC, an imec Research Group, KU Leuven. Her research interests include multi-label, multi-target, hierarchical prediction, recommender systems, tree ensemble learning, survival analysis, and biological network mining.