Advancing Pediatric Heart Transplantation

Developing Predictive Models for Donor Selection

ABSTRACT

Current donor selection practices for pediatric heart transplantation often rely on subjective criteria and lack data-driven precision. To address this gap, we developed machine learning classification models using the Organ Procurement and Transplantation Network Collaborative Histories and Data (ORCHID) dataset, which includes 133,101 deceased donor referrals and 8,972 organ donations across 13 U.S. states. The dataset provides comprehensive, de-identified information on patient demographics, cause of death, referral processes, and organ recovery outcomes. Key features correlated with successful heart transplantation included recovery of other organs (e.g., liver and kidneys), brain death diagnosis, and referrals for multiple tissues. In contrast, death by natural or cardiovascular causes and non-urine output were negatively associated with transplant likelihood.

Statistical analysis revealed brain death status as the strongest discriminator: 99.4% of non-transplanted donors were not brain-dead, while 51.6% of transplanted donors were ($p < 1 \times 10^{-107}$). Tissue and eye referrals were also significantly more common among transplanted donors, suggesting that broader referral engagement may improve heart procurement rates. Microbiology testing, particularly right bronchial Gram stain positivity, occurred more frequently in transplanted donors, which may indicate more thorough evaluation rather than higher infection rates. Modest but significant differences in donor age, height, and weight were observed between groups.

Using five-fold cross-validation, our classification models achieved a mean precision of 0.753 (± 0.065), recall of 0.907 (± 0.036), F1-score of 0.820 (± 0.039), and accuracy of 0.869 (± 0.034). While logistic regression and random forest models showed comparable performance, the ensemble nature of random forests better captured non-linear relationships in the data. These findings highlight the potential for machine learning to enhance pediatric heart donor selection, offering more consistent, evidence-based criteria to support clinical decision-making.

1. Introduction / Background

Pediatric heart transplantation is a life-saving intervention for children with end-stage heart failure, yet the process of donor selection remains largely subjective and lacks standardized, data-driven tools. Given the scarcity of pediatric donor hearts and the critical nature of timely decision-making, improving donor selection is essential for optimizing transplant outcomes and reducing organ discard rates. While clinical judgment plays an important role, the increasing availability of large, structured donor datasets presents an opportunity to enhance selection protocols through machine learning. This project leverages the ORCHID dataset, a comprehensive, multi-center registry of over 133,000 deceased donor referrals and detailed clinical variables, to develop predictive classification models aimed at improving donor selection for pediatric heart transplantation.

2. Related Work

Pediatric heart transplantation remains a critical treatment for children with end-stage heart failure, but the donor selection process is often limited by subjective judgment and lacks standardized, data-driven methodologies. Clinical decision-making frequently hinges on factors such as donor age, ischemia time, infection status, and size matching, yet these variables are weighed inconsistently across transplant centers.[1-3] Moreover, ethical considerations, such as the role of donor obesity, intellectual disability, and evolving debates around xenotransplantation, complicate allocation frameworks.[4-7] As a result, potentially suitable organs may be discarded while other marginal grafts are accepted under pressure, highlighting the need for improved risk stratification and evidence-based donor evaluation.

Efforts to refine donor selection have increasingly turned to outcome predictors such as pre-transplant mechanical support, allosensitization, and diagnosis-specific risk, which are known to impact survival and recovery.[8-10] Registry data also reveal substantial disparities in donor utilization across regions and age groups, with global trends showing rising donor age and variable waitlist mortality.[1] These issues underscore the necessity of tools that can integrate diverse clinical variables to support donor acceptance decisions. Recent clinical innovations, including xenotransplantation trials and algorithmic approaches to high-risk donor assessment, point toward a future in which machine learning may offer scalable and objective solutions.[9] In this context, our

project applies predictive modeling techniques to the ORCHID dataset to enhance precision in pediatric donor selection, with the goal of improving both utilization and post-transplant outcomes.

3. Proposed Work

3.1 Pre-Processing

3.12 Missing data

Columns that had more than 50% missing data were removed. The data was split into training and test sets, and the training dataset was imputed using KNN (K=5). The training set was imputed first and then transferred to the testing data set to prevent data leakage.

3.13 Categorical Variables

Categorical variables were converted to 0 and 1 using one-hot encoding.

3.14 Time Variables

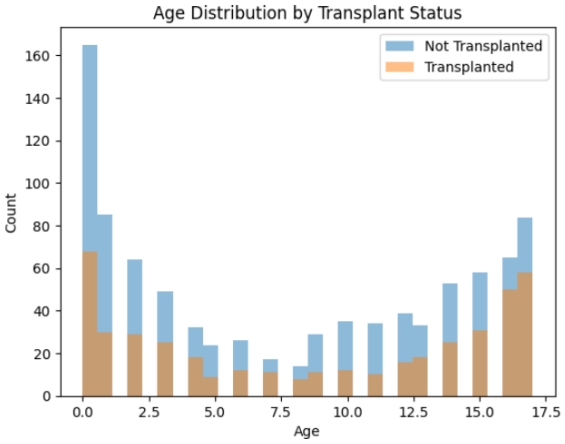
Time variables were parsed accordingly: approached to transplant, approached to authorization, etc

3.2 Exploratory Data Analysis

The dataset was split into transplanted vs non-transplanted. Counts and percentages were obtained for the categorial variables. SD, IQR were obtained for the continuous variables.

Across the heart-donor referrals in this cohort, several binary referral-and-testing flags differ markedly between recipients who ultimately underwent transplantation (group 1) versus those who did not (group 0). Notably, brain death status shows a dramatic shift: almost all non-transplanted donors were not brain-dead (99.4 % vs. 0.6 %), whereas over half of transplanted donors were (51.6 % vs. 48.4 %), $p < 1 \times 10^{-107}$. Referral type also matters: tissue and eye referrals were both more common among transplanted donors (tissue 34.9 % vs. 22.6 %, $p = 3 \times 10^{-4}$; eye 35.4 % vs. 22.7 %, $p = 8 \times 10^{-5}$), suggesting that centers with multi-tissue interests may more aggressively pursue hearts. Several microbiology flags (e.g. positive Gram-stains of the right and left bronchi, sputum) were also significantly more frequent in transplanted cases (for example, R Bronch Gm St: 73 % vs. 50 %, $p = 6 \times 10^{-5}$), which may reflect more intensive peri-donation evaluation rather than a barrier to procurement.

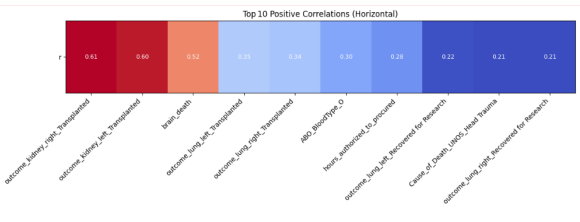
Age of the donor differed only slightly (mean 7.7 years non-transplanted vs. 8.7 years transplanted, $p \approx 0.005$), while Height and Weight were modestly higher in transplanted donors (Height 48.4 in vs. 50.95 in, $p \approx 0.03$; Weight 34.8 kg vs. 41.5 kg, $p \approx 1 \times 10^{-4}$). Timing intervals from brain death or asystole to referral, approach, authorization, and procurement showed no consistent directional differences (all $p > 0.05$), indicating that logistical delays were similar regardless of ultimate transplant status.



Overall, the strongest discriminators of transplantation were clinical donor characteristics (especially brain-death status) and the engagement of multiple referral streams, rather than differences in donor size or process timing.

3.21 Correlations

The correlation heatmap highlights the top 10 features most positively associated with heart transplantation outcomes. The strongest correlations were observed with the transplantation of other organs, specifically the right and left kidneys (r = 0.61 and 0.60, respectively), suggesting that multi-organ viability is a strong indicator of successful heart transplant candidacy. Brain death (r = 0.52) also showed a strong positive correlation, reinforcing its known clinical importance in donor suitability. Moderate correlations were found with lung transplantation outcomes (r ≈ 0.34–0.35), indicating that thoracic organ recovery often occurs in tandem. Additional positive associations included blood type O (r = 0.30), longer authorization-to-procurement times (r = 0.28), and certain research-related lung recoveries and head trauma as cause of death (r = 0.21–0.22). These findings support the hypothesis that comprehensive donor evaluation, optimal physiological condition, and specific donor characteristics contribute meaningfully to heart transplant decisions.



3.3 Modeling

To develop predictive models for pediatric heart donor transplantation, we implemented a structured pipeline comprising feature selection, model development, evaluation, and tuning. Feature selection was performed using Recursive Feature Elimination (RFE), initially with logistic regression and later with support vector machines (SVM) and random forest (RF) classifiers. This process identified the top 20 predictive features, enabling a refined input set for subsequent modeling efforts.

Three classification models were developed and compared: logistic regression, SVM, and random forest. We began with logistic regression using L2 regularization (C=0.1), resolving early issues with multicollinearity that caused matrix singularity errors. The model achieved 91% test accuracy and 5-fold cross-validation (CV) accuracy of 0.883 ± 0.023, providing a stable and interpretable baseline.

Next, we trained a support vector classifier with a radial basis function (RBF) kernel. However, for compatibility with RFE, the kernel was adjusted to linear. We tested several regularization parameters (C = 0.1, 1.0, 10.0), finding optimal performance at C = 0.1 or 1.0, yielding CV accuracy of 0.860 ± 0.031 and test accuracy of 0.89.

We then implemented a Random Forest classifier, capitalizing on its ensemble learning capabilities. We tuned hyperparameters via grid search, testing combinations of n_estimators (50, 100, 200), max_depth (10, 20, None), and min_samples_split (2, 5, 10). The best configuration (n_estimators = 100, max_depth = 10, min_samples_split = 10) achieved a test set accuracy of 0.91 and CV accuracy of 0.869 ± 0.034. RF's architecture leverages bootstrap aggregation (bagging), random feature selection at each node, and majority voting across decision trees. Feature importance was derived from impurity reduction across splits and integrated into the RFE selection process.

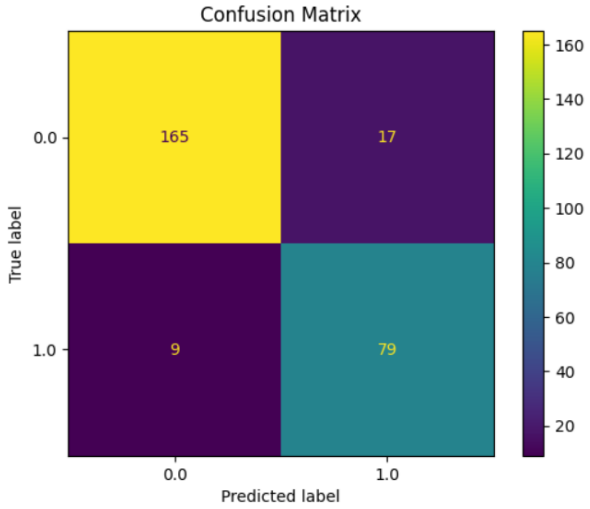
All models were evaluated using precision, recall, F1-score, and accuracy via both test set classification reports and 5-fold cross-validation. Logistic regression yielded the highest overall CV accuracy, while Random Forest outperformed in recall (0.907), making it better suited for identifying positive transplant outcomes. While RF showed slightly lower precision (0.753) and greater variability across folds, its ability to capture non-linear patterns contributed to strong test performance and robustness against overfitting.

Throughout model development, we addressed practical issues including compatibility errors (e.g., SVM kernel requirements for RFE), missing imports, and hyperparameter constraints. Model performance metrics and cross-validation stability informed the final selection, with Random Forest emerging as the most balanced approach in terms of predictive strength and generalizability for transplant prediction. These models provide a framework for improving donor selection by offering data-driven support to clinical decision-making.

4. Evaluation

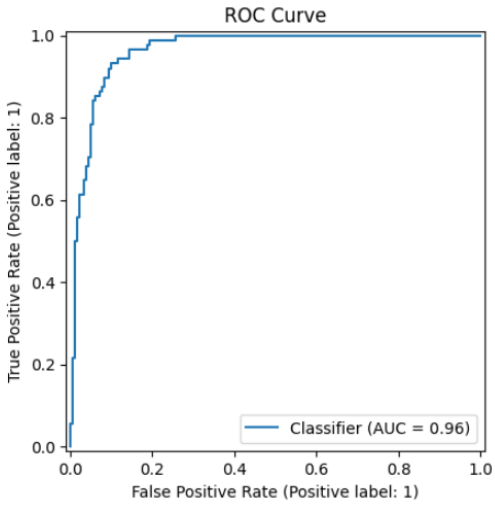
Confusion Matrix

The first plot to help with evaluation was the confusion matrix. The model demonstrates strong performance, achieving high accuracy (90.4%), recall (89.8%), and precision (82.3%). The false negative rate is relatively low, with only 9 misclassified cases out of 88 actual positives, indicating effective identification of true positive instances. However, the model appears to slightly favor the negative class, likely due to the class imbalance reflected in the higher number of true negatives compared to true positives.



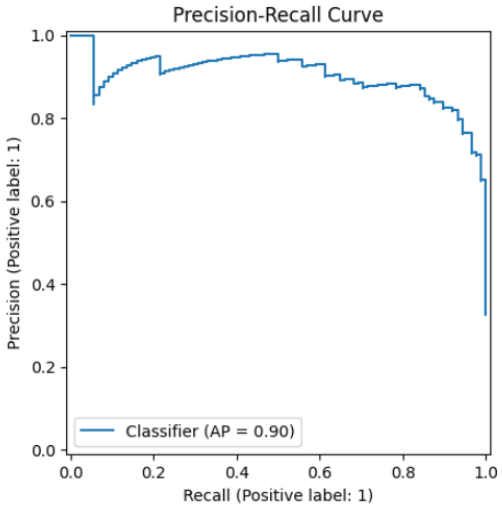
Receive Operating Curve

The ROC (Receiver Operating Characteristic) curve illustrates the trade-off between the true positive rate (sensitivity) and the false positive rate at various classification thresholds. In this case, the ROC curve rises sharply and hugs the top-left corner of the plot, indicating excellent model discrimination between positive (transplanted) and negative (non-transplanted) classes. The area under the curve (AUC) is 0.96, which suggests that the model has a 96% chance of correctly distinguishing a randomly chosen positive case from a negative one. This high AUC value reflects strong overall classification performance, with both high sensitivity and specificity.



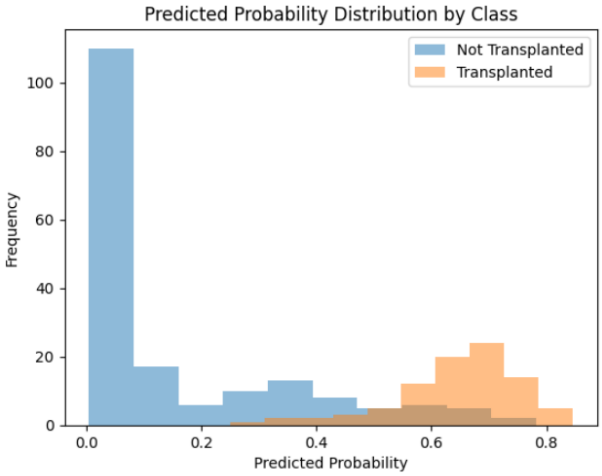
Precision Recall Curve

The Precision-Recall (PR) curve provides a focused evaluation of the model's ability to correctly identify positive cases, particularly valuable in imbalanced datasets. In this case, the curve maintains high precision across a wide range of recall values, indicating that the model consistently makes accurate positive predictions even as it captures more true positives. The average precision (AP) score of 0.90 reflects strong performance, suggesting that the model balances precision and recall effectively. This means it is well-suited for identifying transplanted donors with minimal false positives, making it a valuable tool in clinical decision-making where the cost of misclassification can be high.



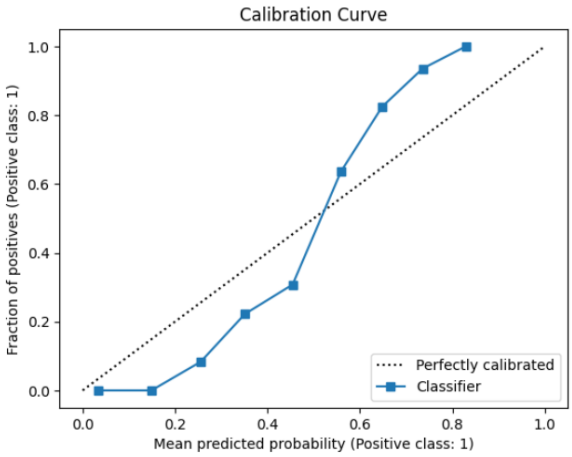
Predicted Probabilities Distribution Plot

The predicted probability distribution plot shows clear separation between the two outcome classes, with non-transplanted donors clustered near lower predicted probabilities and transplanted donors skewed toward higher values. This indicates that the model assigns probabilities in a meaningful and discriminative way, effectively distinguishing between the classes. While some overlap exists—suggesting a few ambiguous cases—the overall distribution supports the model's confidence in most predictions. The concentration of class 0 predictions near 0 and class 1 predictions near 1 demonstrates that the model produces well-calibrated probabilities, which is critical for decision-making thresholds in clinical settings.



Calibration Plot

The calibration plot assesses how well the predicted probabilities align with actual outcomes. In this case, the calibration curve closely follows the diagonal reference line at higher probability ranges, indicating that the model is well-calibrated for confident predictions. However, at lower probability levels, the curve dips slightly below the diagonal, suggesting that the model may be somewhat underconfident in these lower-probability predictions. Overall, the model demonstrates good calibration, especially in the clinically relevant higher-probability ranges, meaning that its predicted probabilities can be reliably interpreted as true likelihoods—an important consideration for risk-based clinical decision-making.



5. Discussion

The classification models developed in this study—particularly the Random Forest and Logistic Regression classifiers—demonstrated strong overall performance in predicting heart transplant outcomes from the ORCHID dataset. With a test accuracy of 91% and an average cross-validation accuracy of 86.9–88.3%, the models consistently identified relevant patterns in donor data. Notably, the Random Forest model achieved a high recall of 90.7%, suggesting excellent sensitivity in detecting actual transplant cases. This makes it especially valuable in clinical settings where false negatives (missed transplant opportunities) could lead to suboptimal organ utilization.

The ROC curve revealed a strong ability to discriminate between transplanted and non-transplanted donors, with an area under the curve (AUC) of 0.96. This indicates that the model has a 96% chance of correctly distinguishing between a randomly selected positive and negative case. The Precision-Recall (PR) curve further supports the model's performance, with an average precision (AP) of 0.90. The high PR performance, particularly in the context of class imbalance, suggests that the model is effective at maintaining precision while achieving high recall, a key requirement when the positive class (transplanted donors) is less common.

The predicted probability distribution plot showed clear separation between the classes, with non-transplanted donors centered near low predicted probabilities and transplanted donors skewed toward high probabilities. This distribution indicates that the model outputs probabilities that reflect actual class distinctions, enabling confident interpretation of risk scores. The calibration plot reinforces this finding: predicted probabilities at higher levels closely matched observed outcomes, indicating that the model is well-calibrated in clinically actionable regions. Minor underconfidence at lower predicted probabilities was noted but is less critical given the clinical focus on higher-risk cases.

These findings suggest that data-driven modeling can enhance the precision of pediatric donor selection by providing objective risk assessments based on real-world referral data. In particular, the high recall and well-calibrated probability outputs of the Random Forest model support its use as a decision-support tool to identify suitable donors who may otherwise be overlooked. The ability to flag high-risk candidates with reliable probability estimates could improve organ utilization and ultimately expand access to transplantation for pediatric patients.

Limitations

While the models showed strong performance metrics, there are limitations to consider. Some degree of class imbalance remained, and although mitigated through evaluation metrics like the PR curve, it may still influence prediction bias. Furthermore, the generalizability of the model outside the 13-state ORCHID dataset remains untested.

6. Conclusion

This study demonstrates the potential of machine learning models, particularly Random Forest and Logistic Regression, to improve pediatric heart donor selection using the ORCHID dataset. The models achieved strong performance, with high accuracy, recall, and well-calibrated predicted probabilities, offering a data-driven approach to support clinical decision-making. By identifying key predictors such as brain death status, organ recovery patterns, and referral types, the models provide meaningful insights into the donor selection process. These tools can enhance transplant efficiency and reduce missed opportunities for eligible donors.

Future work

Future work should focus on external validation across broader geographic regions, integration of recipient outcomes to better align donor-recipient matching, and exploration of time-dependent features to enable dynamic, real-time predictions. Enhancing model calibration and incorporating clinical decision curve analysis will also be critical to optimizing the model's utility in practice.

Timeline

The project followed a structured five-week timeline, progressing from data preparation to final documentation. In Week 1, the focus was on data filtering and initial preprocessing to prepare the dataset for analysis. Week 2 involved completing the preprocessing pipeline, including handling missing data and applying one-hot encoding. Additionally, exploratory data analysis was conducted using methods such as chi-square tests and box plots to identify patterns and outliers. During Weeks 3 and 4, various machine learning models were explored and iteratively refined, culminating in the evaluation of the final selected model. Finally, in Week 5, the

project was wrapped up with the completion of the final report and synthesis of results.

ACKNOWLEDGMENTS

None

REFERENCES

1. Elizer S, Mantell BS. Risk factors affecting mortality in pediatric heart transplantation: A comprehensive review of pre- and post-transplant contributors. JHLT Open. 2025;9:100309.
2. Conway J, Ballweg JA, Fenton M, Kindel S, Chrisant M, Weintraub RG, et al. Review of the impact of donor characteristics on pediatric heart transplant outcomes. Pediatr Transplant. 2020;24(3):e13680.
3. Donne M, De Pauw M, Vandekerckhove K, Bove T, Panzer J. Ethical and practical dilemmas in cardiac transplantation in infants: a literature review. Eur J Pediatr. 2021;180(8):2359-65.
4. Wilkens SJ, Gossett JG, Patel A. Ethical Considerations in Pediatric Heart Transplantation. In: Mavroudis C, Cook JT, Mavroudis CD, editors. Bioethical Controversies in Pediatric Cardiology and Cardiac Surgery. Cham: Springer International Publishing; 2020. p. 217-35.
5. Berkman E, Wightman A, Friedland-Little JM, Albers EL, Diekema D, Lewis-Newby M. An ethical analysis of obesity as a determinant of pediatric heart transplant candidacy. Pediatr Transplant. 2021;25(3):e13913.
6. Hurst DJ, Padilla L, Merlocco A, Rodger D, Bobier C, Gray WH, et al. Pediatric Cardiac Xenotransplantation: Recommendations for the Ethical Design of Clinical Trials. Transplantation. 2024;108(10):e292-e300.
7. Singh TP, Hsich E, Cherikh WS, Chambers DC, Harhay MO, Hayes D, Jr., et al. The International Thoracic Organ Transplant Registry of the International Society for Heart and Lung Transplantation: 23rd pediatric heart transplantation report-2020; focus on deceased donor characteristics. J Heart Lung Transplant. 2020;39(10):1028-37.
8. Singh N, Raees MA, Zafar F. Donor considerations in pediatric heart transplantation. Transl Pediatr. 2019;8(4):284-9.
9. Bearl DW. Ethical issues in access, listing and regulation of pediatric heart transplantation. Transl Pediatr. 2019;8(4):278-83.
10. Canter CE, Shaddy RE, Bernstein D, Hsu DT, Chrisant MR, Kirklin JK, et al. Indications for heart transplantation in pediatric heart disease: a scientific statement from the American Heart Association Council on Cardiovascular Disease in the Young; the Councils on Clinical Cardiology, Cardiovascular Nursing, and Cardiovascular Surgery and Anesthesia; and the Quality of Care and Outcomes Research Interdisciplinary Working Group. Circulation. 2007;115(5):658-76.