Advancing Pediatric Heart Transplantation: Predictive Models for Donor Selection

DTSA 5506: DATA MINING PROJECT

Problem Statement

- Current donor selection for pediatric heart transplantation lacks data-driven precision

- This project aims to develop classification models to improve donor selection

Related Work

Focus Area	Notable Issues	Key Sources
Donor Selection	Age, ischemia time, infection status, weight matching	Singh et al. (2020); Singh et al. (2019); Conway et al. (2020)
Ethical Considerations	Obesity, intellectual disability, resource allocation, xenotransplantation	Berkman & Wightman (2021); Hurst et al. (2024); Wilkens et al. (2020); Bearl (2019)
Outcome Predictors	Pre-transplant support, diagnosis, panel reactive antibody, BMI	Mantell & Elizer (2025); Donné et al. (2021); Canter et al. (2007)
Registry & Trends	Global disparities, median donor age, mortality trends	Singh et al. (2020); AHA Guidelines (2007)
Clinical Innovations	Xenotransplantation trials, high-risk acceptance algorithms	Hurst et al. (2024); Nabzdyk et al. (2024)

Proposed Work

- Dataset: https://physionet.org/content/orchid/2.0.0/
- ORCHID Dataset Overview
  - 133,101 deceased donor referrals
  - 8,972 organ donations
  - 13 states covered
- Referrals by OPO
  - OPO 1: 32,148 potential donors
  - OPO 2: 16,144 potential donors
  - OPO 3: 12,516 potential donors
  - OPO 4: 33,641 potential donors
  - OPO 5: 15,738 potential donors
  - OPO 6: 22,914 potential donors

Proposed Work

▶ All data was de-identified in accordance with HIPAA standards using structured data cleansing and date shifting.ORCHID consists of ten tables (CSV files), linked via PatientID, and grouped into three categories:

▶ OPO Referrals

  ▶ Referral information: patient demographics & cause of death

  ▶ Process data: timestamps for every action (next-of-kin approach, authorization, procurement, death modes)

  ▶ Outcomes: binary flags for approached, authorized, procured; plus per-organ recovery results

Proposed Work

- OPO Events
  - ChemistryEvents: blood chemistry (kidney panel, LFTs, electrolytes)
  - CBCEvents: complete blood count with differential
  - ABGEvents: arterial blood gas + ventilator settings
  - SerologyEvents: presence/absence of donation-relevant antigens/antibodies
  - CultureEvents: infection culture results (blood, urine, other)
  - HemoEvents: hemodynamics over time_event_start → time_event_end (or point measurements)
  - FluidBalanceEvents: fluid intake/output over time_event_start → time_event_end

- OPO Deaths Captures referring-hospital death data (asystole, brain death timestamps, etc.)

Proposed Work

- Preprocessing
  - Evaluate missing data
  - One hot encoding
- Exploratory Data Analysis
  - Evaluate for anomalies
  - Chi Square
  - Correlations
- Modeling
  - Supervised ML for classifcation
    - Consider XGboost
    - Consider Linear Regression
    - Consider SVM
    - Consider RandomForest

Evaluation

- Confusion matrix
- Accuracy
- F1 score
- Precision
- Recall
- Plots: precision-recall, ROC, calibration, test/validation, learning curve, K fold validation (k=5)

Timeline

- **Week 1**
  - Data filtering, initial data preprocessing
- **Week 2**
  - Finish preprocessing including missing data and one hot encoding
  - Exploratory data analysis including correlctions, chi square, box plots, etc
- **Week 3-4**
  - Explore machine learning models
  - Evaluation of final model
- **Week 5**
  - Complete final write-up