

Bridging the Bias Gap: A Human-Centered Approach to Fair Post-Transplant Survival Prediction Across Racial Groups

Introduction:

ML models are used in healthcare to predict outcomes like cancer detection, survival rate and drug efficacy. Current predictive models often fall short in addressing disparities related to socioeconomic status, race, and geography. Addressing these gaps is crucial for enhancing patient care, optimizing resource utilization, and rebuilding trust in the healthcare system. These biased predictions are propagated due to imbalanced datasets and incorrect algorithms. This project tries to tackle this critical challenge in human-centered AI.

Problem statement:

Develop ML models to improve the prediction of transplant survival rates for patients undergoing allogeneic Hematopoietic Cell Transplantation (HCT) regardless of their ethnic and racial background.

Motivation:

I felt this is an important topic because I can potentially learn how to work on biases in medical dataset across different demographic groups and to develop models that provide fair predictions for the same. The goal is to address disparities by bridging diverse data sources, refining algorithms, and reducing biases to ensure equitable outcomes for patients across diverse race groups. Creating interpretable predictions that healthcare providers can trust across demographic groups is the main challenge of this project. Data analysis for bias, understanding fairness metrics, model development and fairness optimization will be the phases of the proposed project.

Dataset:

Tushar Deshpande, Deniz Akdemir, Walter Reade, Ashley Chow, Maggie Demkin, and Yung-Tsi Bolon. CIBMTR - Equity in post-HCT Survival Predictions.

<https://kaggle.com/competitions/equity-post-HCT-survival-predictions>, 2024. Kaggle.

Dataset description:

I felt this dataset is perfect because:

- a. It encompasses a range of demographic and medical characteristics of both recipients and donors, such as age, sex, ethnicity, disease status, and treatment details.
- b. It features equal representation across recipient racial categories including White, Asian, African American, Native American, Pacific Islander, and More than One Race.
- c. There are no privacy concerns since this dataset is synthetically generated.