Using Reweighting to Reduce Racial Bias in Adverse Event Prediction Models

Amalia Ionescu, PhD, Emily F. Wong, PhD, MS, Priyanka Merchant, MS, Nicole Beyrouthy, MS, RN-BC, Michael Stange, BS, Matthew Moore, MS, Andrew Hudson, MD, Tiffani J. Bright, PhD

Cedars Sinai Medical Center, Los Angeles, CA, USA

Introduction

- Algorithmic bias occurs when systematic errors in machine learning pipelines skew scientific study results, potentially leading to inaccurate or unjust outcomes.
    - This bias can arise from unrepresentative samples, incomplete data, or flawed study designs.
- Bias in biomedical research can lead to non-generalizable findings, resulting in suboptimal or even harmful medical treatment for certain groups.
    - Previous work found that a widely used healthcare algorithm underestimated the needs of Black patients compared to White patients (Obermeyer et al., 2019).
- Fairness in biomedical research aims to minimize disparities in healthcare access and treatment efficacy by ensuring diverse and representative study populations.
- One method to mitigate bias is **reweighing**, which assigns different weights to data points in a training set to improve fairness in model predictions, prioritizing historically disadvantaged individuals.
- **Current Study:** Fit and evaluate novel model that utilizes reweighing during training to equitably predict adverse events across racial/ethnic groups in a Cedars-Sinai patient dataset.

Methodology

- Used a dataset obtained from Cedars-Sinai with 81,258 unique patient encounters across three years.
- Dataset is split into training, validation, and test sets.
    - Applied random under-sampling to training set to address class imbalance.
- Fit random forest classifier models (base model vs. weighted model) using race as a sensitive feature.
    - Finetuned the models using multi-objective optimization to both increase the AUROC (Area Under the Receiver Operating Characteristic curve) and decrease the FPR (false positive rate).
- Picked best model from the pareto-front.
- Base and weighted models evaluated against Epic's proprietary model on test datasets.

References and Acknowledgements

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447-453.

Results

- Our finetuned weighted model achieved a balanced accuracy of 0.64, while the Epic model achieved 0.68.
- When divided by race, our model had an overall lower false negative rate (FNR) for each category as opposed to the Epic model (see Figs. 1 and 2), yet higher false positive rates.

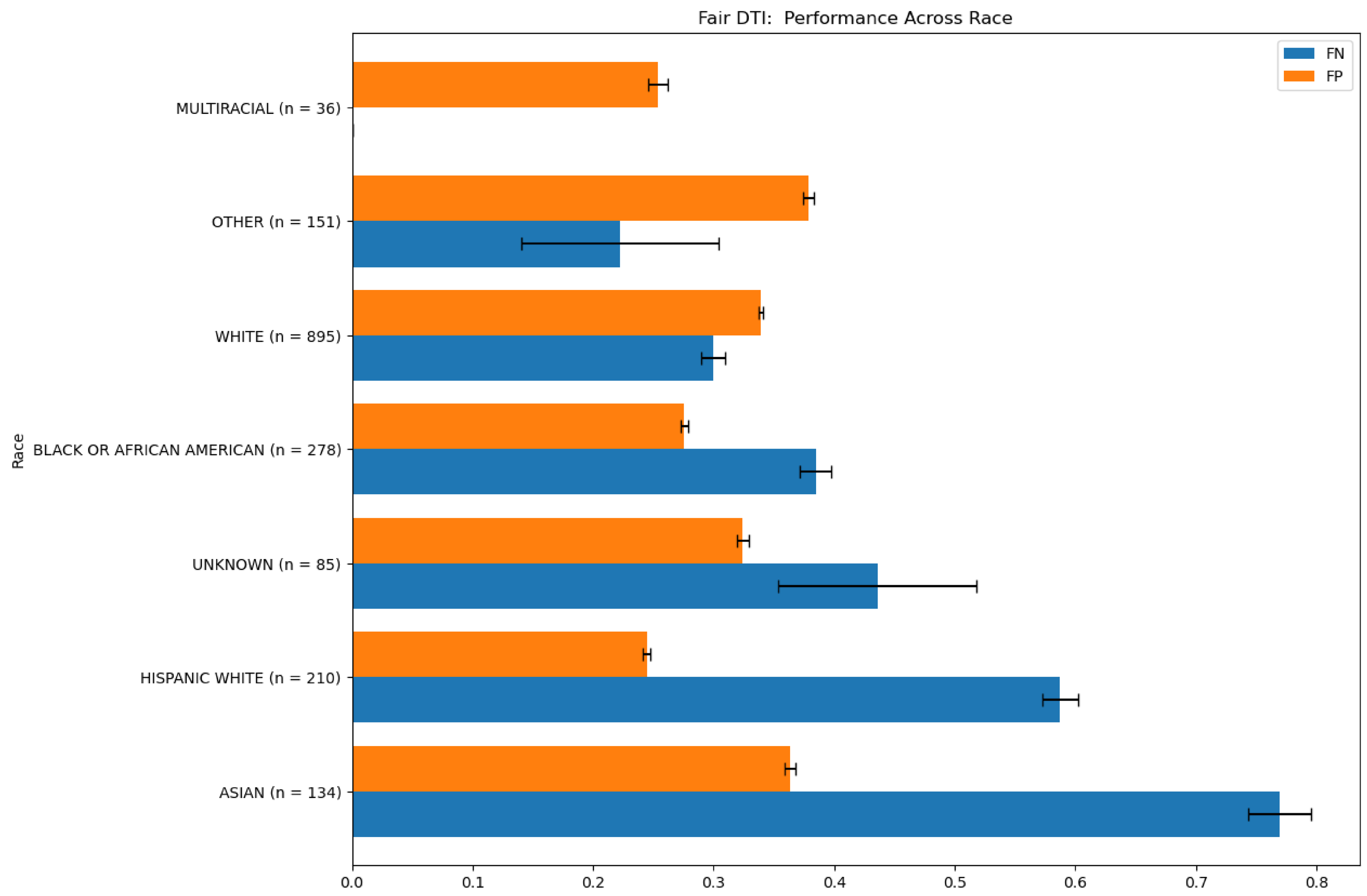

Figure 1. Our weighted model's false negative and false positive rates by race.

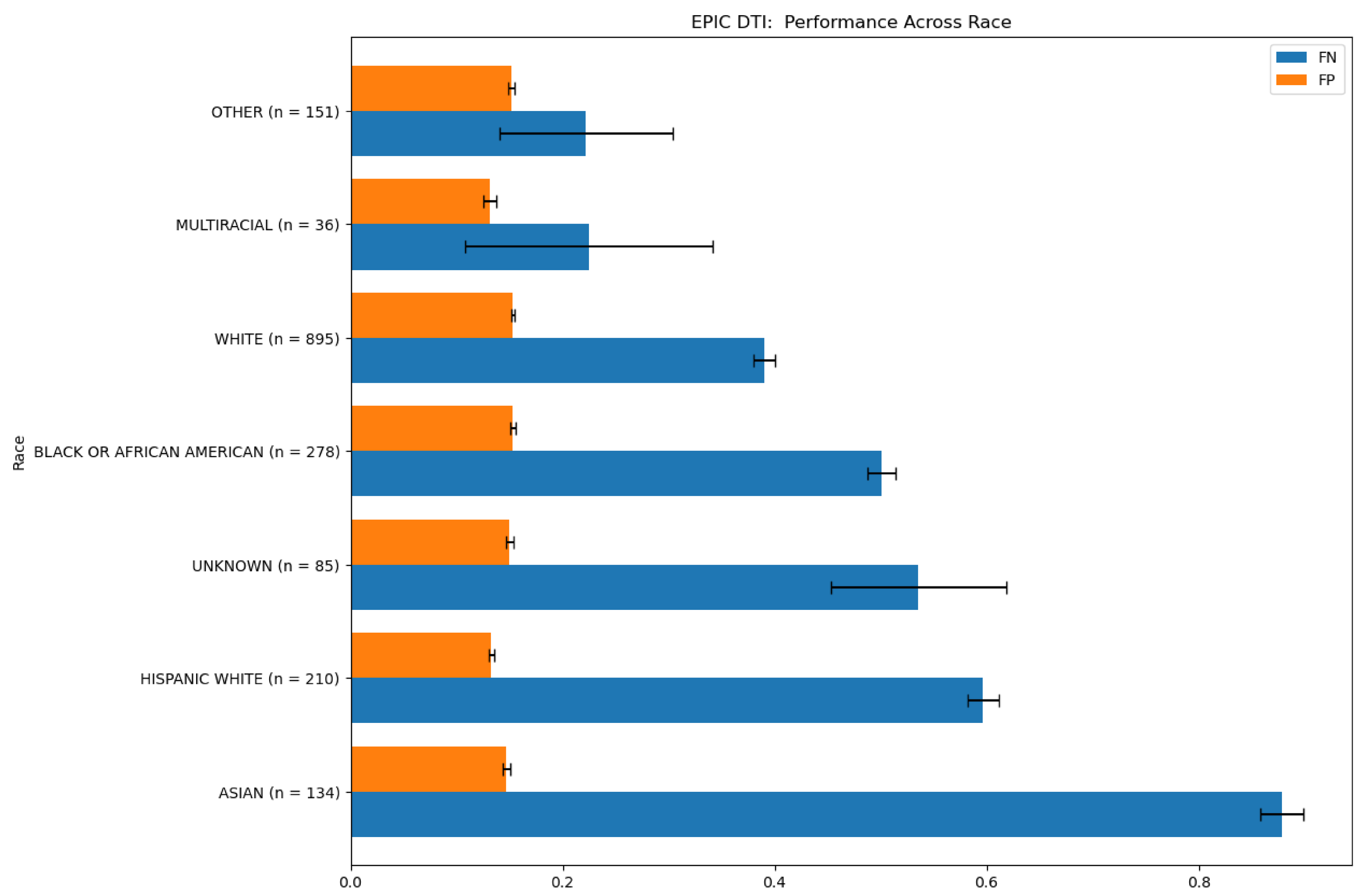

Figure 2. Epic model's false negative and false positive rates by race.

Conclusion

- Our results show that Epic's DTI model systematically underestimates risk of deterioration for Asian patients.
- While reweighting decreases overall FNRs, it may not be sufficient for resolving overall patterns of bias.
- Future work will focus on identifying the features that serve as the most crucial predictors, in order to optimize the accuracy of the model for all groups.

Cedars Sinai