FAIRNESS-CONSTRAINED LOGISTIC REGRESSION ACHIEVES SUPERIOR PERFORMANCE AND RACIAL EQUITY IN BREAST CANCER SURVIVAL PREDICTION

Jack A. Zeineh[1,2], Rebecca E. DeAngel[2], Enrique R. Grullon[2], Thomas J. Lawton[3], and Kenneth J. Bloom[4]

[1] Icahn School of Medicine, [2] ScopeData, [3] Department of Pathology and Laboratory Medicine, UC Irvine, [4] Nucleai

SAN ANTONIO BREAST CANCER SYMPOSIUM

PS1-13-04

Background

Machine learning models for clinical prediction can underperform for African American (AA) patients, potentially delaying diagnosis, leading to inappropriate treatment recommendations, and reducing access to clinical trials — all of which can worsen survival disparities.[1] Standard optimization focuses solely on overall performance, potentially exacerbating such inequities.

Research question

Can fairness-constrained hyperparameter optimization simultaneously improve both overall clinical performance and AA patient outcomes in breast cancer survival prediction?

Methods

Data

- **Dataset**: TCGA-BRCA[2] (N=1,000); 18.6% AA; 70.8% 5-year survival
- **Features**: 50 clinical/genomic variables; IPCW adjustments[3]
- **Validation**: 10-fold stratified CV with matched folds; paired t-tests + bootstrap CIs
- **Train/Test split**: 67% train / 33% test; stratified and matched by race & stage to maintain outcome balance across AA and non-AA groups

Control arm

- Baseline logistic regression
- Objective: maximize overall AUC
- Standard hyperparameter tuning
- No fairness penalties

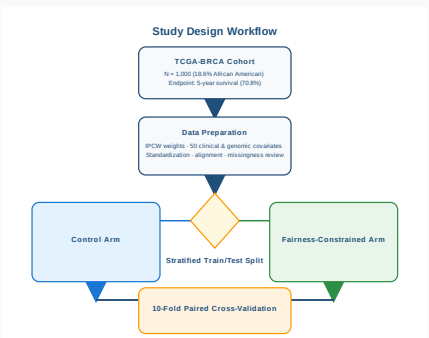


Figure 1 — Study design flowchart: data → preprocessing → matched splits → parallel optimization → statistical comparison

Fairness constrained arm

We performed a multi-objective, fairness-constrained hyperparameter optimization inside each cross-validation fold.[4,5] The procedure used a randomized search over model hyperparameters (for example: L2 regularization strength, class weights, and decision threshold) together with a scalar fairness penalty λ that trades off overall discrimination against subgroup disparity. We defined disparity as the absolute difference between subgroup AUCs (disparity = |AUC_AA − AUC_non-AA|) and used a combined selection score of the form: *score* = *AUC_overall − λ · disparity*. By sweeping λ we produced a Pareto frontier of models and selected the fairness-constrained solution that improved AA performance while keeping any overall-AUC loss within a small, pre-specified tolerance.

Cross validation

- 10-fold paired cross-validation
- Matched folds; paired t-tests; bootstrap confidence intervals (n=1,000 resamples)
- Evaluates both overall and subgroup performance
- Endpoints: overall AUC and AA AUC; disparity = $AUC_{non\text{-}AA} - AUC_{AA}$

Results

Overall Performance Improvement

Control (Standard Optimization) — Overall AUC: **0.634** (95% CI: 0.582–0.664)

Control (AA subgroup) — AUC: **0.610** (95% CI: 0.518–0.687)

Treatment (Fairness-Constrained) — Overall AUC: **0.641** (95% CI: 0.592–0.669)

Improvement: **+0.007 AUC** (95% CI: +0.002 to +0.013), p = 0.018, Cohen's d = 0.917

Clinical meaning: modest population-level AUC gains that may reduce misclassification across risk strata and improve average risk stratification.



Figure 2 — Overall performance comparison (Control vs Fairness-Constrained)

African American Performance Enhancement

Control AA AUC: **0.610** (95% CI: 0.518–0.687)

Fairness AA AUC: **0.632** (95% CI: 0.528–0.713)

Improvement: **+0.021 AUC** (95% CI: +0.003 to +0.040), p = 0.027, Cohen's d = 0.832

Clinical meaning: improved discrimination in AA patients—likely fewer missed high-risk AA patients and more equitable risk guidance.

Non-AA Performance Maintained

Non-AA AUC: **0.639** (95% CI: 0.596–0.678); no significant change with fairness constraints (p = 0.276)

Clinical meaning: no detectable compromise to majority-group prediction performance.

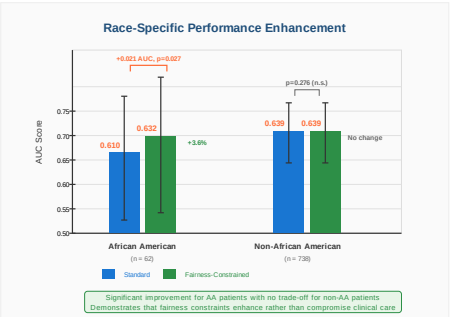


Figure 3 — Race-specific AUC by group and model

Equity Achievement

Control disparity: +0.029 AUC (favoring non-AA). Fairness disparity: -0.012 AUC (slight AA advantage). Net improvement: 41-point reduction in disparity.

Clinical Significance

- **Dual benefit:** Improved overall discrimination and AA-specific prediction.
- **Statistical robustness:** Improvements meet conventional thresholds (p < 0.05) and were confirmed by bootstrap resampling (n = 1,000).
- **Large effect sizes:** Cohen's d > 0.8 for main comparisons indicates meaningful clinical effect.
- **No detected trade-off:** Non-AA performance unchanged, supporting equity without harming majority-group accuracy.

Practical implication: These changes are likely to reduce misclassification in AA patients and improve individualized treatment discussions; further external validation is recommended before deployment.

Cross-Validation Robustness

10-fold CV: consistent improvements across folds and minimal variance for the fairness-constrained approach.

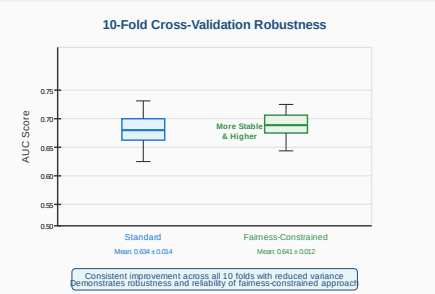


Figure 4 — Cross-validation stability: AUC distributions across folds

Implementation notes and recommended next steps:

1. External validation on independent cohorts is required before deployment to ensure transportability across clinical settings.
2. Report both discrimination and calibration by subgroup; decision thresholds should be evaluated for clinical utility (net benefit) rather than AUC alone.
3. Operational monitoring: include continuous fairness and performance checks (subgroup AUCs, calibration, and changes in case-mix) following deployment.
4. Where feasible, consider prospective evaluation to measure whether improved predictions lead to measurable changes in clinical decisions or outcomes.

Conclusions

Fairness-constrained hyperparameter optimization achieved dual improvements in clinical performance and racial equity for breast cancer survival prediction. Equitable ML can be clinically superior.

Key Findings

- **Overall AUC:** 0.634 → 0.641 (+0.007; 95% CI +0.002 to +0.013; p=0.018)
- **AA AUC:** 0.610 → 0.632 (+0.021; 95% CI +0.003 to +0.040; p=0.027)
- **Disparity:** +0.029 → −0.012 (41-point improvement)
- **Non-AA AUC:** 0.639 (no significant change; p=0.276)

Clinical take-home: Fairness-constrained training can improve AA-specific prediction without compromising non-AA accuracy.

References

1. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366(6464):447–453. doi: 10.1126/science.aax2342
2. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature. 2012;490(7418):61–70. doi: 10.1038/nature11412
3. Cole SR, Hernán MA. Adjusted survival curves with inverse probability weights. Comput Methods Programs Biomed. 2004;75(1):45–49. doi: 10.1016/j.cmpb.2003.10.004
4. Hardt M, Price E, Srebro N. Equality of Opportunity in Supervised Learning. arXiv:1610.02413 (2016). arXiv
5. Rezaei A, Fathony R, Memarrast O, Ziebart B. Fairness for Robust Log Loss Classification. AAAI (2020). arXiv/AAAI

Contact

Corresponding author: jackzemail@gmail.com

Copies of this poster obtained through Quick Response (QR) Code are for personal use only and may not be reproduced without permission from SABCS® and the author.