USC Mann
Alfred E. Mann School of Pharmacy
and Pharmaceutical Sciences

Evaluating Fairness Across Machine Learning Algorithms in Health Models Incorporating Race/Ethnicity as Predictors

Yizhi Liang, MS
Mann School of Pharmacy and Pharmaceutical Sciences
University of Southern California

Acknowledgment

- This study was a pure student-involved modeling practice from a random talk during lunchtime with Dr. Chen Sheng (Cardiologist).
- Contributions:
  - Chen Sheng and Yizhi Liang conceptualized the study
  - Beier Chen collected the data
  - Jize Luo and Yizhi Liang built the model
  - Chen Sheng checked its clinical relevance

Key Info

Like many economists said, to answer a question:

Should we include race in the machine learning models?

The answer is always:

IT DEPENDS ON CONTEXTS

- Evidence is mixed
- A guideline for model-building practice is needed

Outline

- Background for Machine Learning and Racial Fairness Question
- Data Sources and Model Building Tricks
- Performance Comparisons
- Lessons from Practice

Outline

- **Background for Machine Learning and Racial Fairness Question**
- Data Sources and Model Building Tricks
- Performance Comparisons
- Lessons from Practice

USCMann
Alfred E. Mann School of Pharmacy and Pharmaceutical Sciences

The use of race in models is still contentious


PNAS RESEARCH ARTICLE | ECONOMIC SCIENCES MEDICAL SCIENCES

Check for updates

Using measures of race to make clinical predictions: Decision making, patient health, and fairness

Charles F. Manski[a,b,1], John Mullahy[c], and Atheendar S. Venkataramani[d]

Contributed by Charles F. Manski; received February 27, 2023; accepted May 24, 2023; reviewed by Marcella Alsan and Ziad Obermeyer

The use of race measures in clinical prediction models is contentious. We seek to inform the discourse by evaluating the inclusion of race in probabilistic predictions of illness that support clinical decision making. Adopting a static utilitarian framework to formalize social welfare, we show that patients of all races benefit when clinical decisions are jointly guided by patient race and other observable covariates. Similar conclusions emerge when the model is extended to a two-period setting where prevention activities target systemic drivers of disease. We also discuss non-utilitarian concepts that have been proposed to guide allocation of health care resources.

clinical prediction | patient care | utilitarian welfare analysis | race

Significance

The use of race measures in clinical prediction models is contentious. We seek to inform the discourse by evaluating the inclusion of race in probabilistic predictions of illness that support clinical decision making. We

- All patients benefit when clinical decisions are jointly guided by race and other covariates
- In prognostic models, the use of race will compromise equality because of an ex ante rewarding
- The choice of convenient, seemingly effective proxies for health outcomes is the source of algorithmic bias

The use of race in models is still contentious

SCIENCE ADVANCES | RESEARCH ARTICLE

HEALTH AND MEDICINE

Use of race in clinical algorithms

Anirban Basu[1,2]*

To answer whether patients' race belongs in clinical prediction algorithms, two types of prediction models are considered: (i) diagnostic, which describes a patient's clinical characteristics, and (ii) prognostic, which forecasts a clinical risk or treatment effect that a patient is likely to experience in the future. The ex ante equality of opportunity framework is used, where specific health outcomes, which are prediction targets, evolve dynamically due to the effects of legacy levels of outcomes, circumstances, and current individual efforts. In practical settings, this study shows that failure to include race corrections will propagate systemic inequities and discrimination in any diagnostic model and specific prognostic models that inform decisions by invoking an ex ante compensation principle. In contrast, including race in prognostic models that inform resource allocations following an ex ante reward principle can compromise the equality of opportunities for patients from different races. Simulation results demonstrate these arguments.

Diagnostic prediction: failure to include race will propagate inequities and discrimination

Prognostic prediction: including race with resource allocations following an ex ante reward principle can compromise the equality

- All patients benefit when clinical decisions are jointly guided by race and other covariates
- In prognostic models, the use of race will compromise equality because of an ex ante rewarding
- The choice of convenient, seemingly effective proxies for health outcomes is the source of algorithmic bias

USC Mann
Alfred E. Mann School of Pharmacy and Pharmaceutical Sciences

The use of race in models is still contentious

RESEARCH

RESEARCH ARTICLE

ECONOMICS

Dissecting racial bias in an algorithm used to manage the health of populations

Ziad Obermeyer[1,2]*, Brian Powers[3], Christine Vogeli[4], Sendhil Mullainathan[5]*†

Health systems rely on commercial prediction algorithms to identify and help patients with complex health needs. We show that a widely used algorithm, typical of this industry-wide approach and affecting millions of patients, exhibits significant racial bias: At a given risk score, Black patients are considerably sicker than White patients, as evidenced by signs of uncontrolled illnesses. Remedying this disparity would increase the percentage of Black patients receiving additional help from 17.7 to 46.5%. The bias arises because the algorithm predicts health care costs rather than illness, but unequal access to care means that we spend less money caring for Black patients than for White patients. Thus, despite health care cost appearing to be an effective proxy for health by some measures of predictive accuracy, large racial biases arise. We suggest that the choice of convenient, seemingly effective proxies for ground truth can be an important source of algorithmic bias in many contexts.

- All patients benefit when clinical decisions are jointly guided by race and other covariates
- In prognostic models, the use of race will compromise equality because of an ex ante rewarding
- The choice of convenient, seemingly effective proxies for health outcomes is the source of algorithmic bias

USC Mann
Alfred E. Mann School of Pharmacy and Pharmaceutical Sciences

2023 Novel heart disease risk calculator Specifically, Sex-Specific and Race-Free

Circulation
Volume 149, Issue 6, 6 February 2024; Pages 430-449
https://doi-org.libproxy2.usc.edu/10.1161/CIRCULATIONAHA.123.067626

ORIGINAL RESEARCH ARTICLE

Development and Validation of the American Heart Association's PREVENT Equations

Sadiya S. Khan, MD, MSc, Kunihiro Matsushita, MD, PhD, Yingying Sang, , Shoshana H. Ballew, PhD, Morgan E. Grams, MD, PhD, Aditya Surapane, Michael J. Blaha, MD, MPH, April P. Carson, PhD, Alexander R. Chang MS, Elizabeth Ciemins, MPH, PhD, Alan S. Go, MD, Orlando M. Gutierr, Shih-Jen Hwang, PhD, Simerjot K. Jassal, MD, MAS, Csaba P. Koves Donald M. Lloyd-Jones, MD, ScM, Michael G. Shlipak, MD, MPH, Latha P. Palaniappan, MD, MS, Laurence Sperling, MD, Salim S. Virani, MD, PhD Katherine Tuttle, MD, Ian J. Neeland, MD, Sheryl L. Chow, PharmD, Janan Rangaswami, MD, Michael J. Pencina, PhD, Chiadi E. Ndumele, MD, PhD Josef Coresh, MD, PhD for the Chronic Kidney Disease Prognosis Consortiu the American Heart Association Cardiovascular-Kidney-Metabolic Science Advis Group*

AHA PREVENT Equations

- Remove race from risk prediction, acknowledging it is a social construct and not a biological predictor
- Build sex-specific models, acknowledging biological differences by sex
- Include a measure of place-based social disadvantage support

Outline

- Background for Machine Learning and Racial Fairness Question
- **Data Sources and Model Building Tricks**
- Performance Comparisons
- Lessons from Practice

Two Tasks & Two Data & Two Scenarios

Task 1: CVD Prediction

- Data Source: National Health and Nutrition Examination Surveys (NHANES) 2007-2018
- Binary outcome:
  - Self-report cardiovascular diseases, including coronary heart disease, heart failure, or stroke
- Models:
  - All population model
  - Race-specific model
  - Gender-specific model

Task 2: APO Prediction

- Data Source: U.S. Birth Registration 2016-2023
- Binary outcome
  - Adverse pregnancy outcomes, including cesarean delivery, ICU admission, transfusion, preterm birth, low birthweight, or NICU admission
- Models:
  - All population model
  - Race-specific model

Scenarios: (1) Race-Sensitive (RS), including race as a predictor; (2) Race-Neutral (RN)

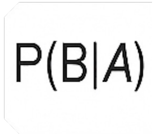
7 Learners: Compared with Logistic Regression

K-Nearest Neighbors (KNN)

Classified based on the majority class of nearest data points

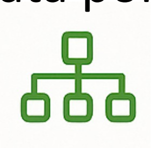
Support Vector Machine (SVM)

Classified based on the best separating hyperplane



Naïve Bayes

Classified based on the Bayes' theorem with feature independence



Decision Tree

Classified based on the splitting features

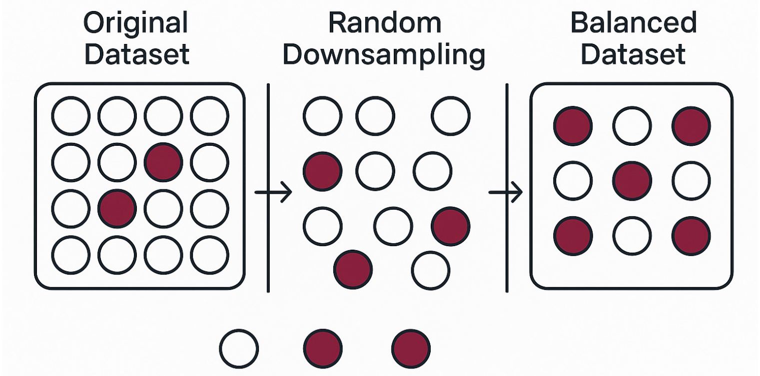
Random Forest

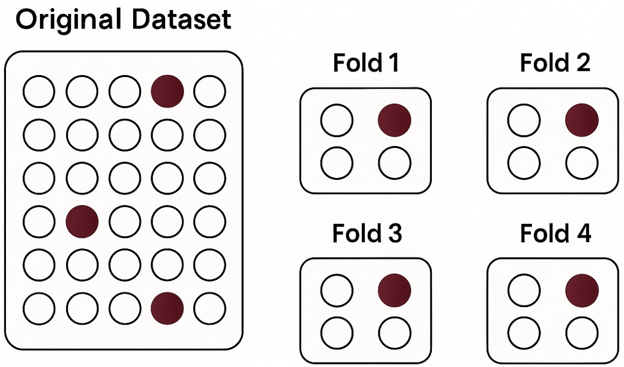
Classified based on the ensemble of decision trees

XGBoost

A gradient boosting method that builds trees sequentially, min. error

Modeling Concern 1: *Imbalanced Data*
Trick: *Downsampling & Stratification*





Downsampling

Remove random samples from the majority class

- Simplify decision boundary

Stratification

Preserve class proportion during cross-validation

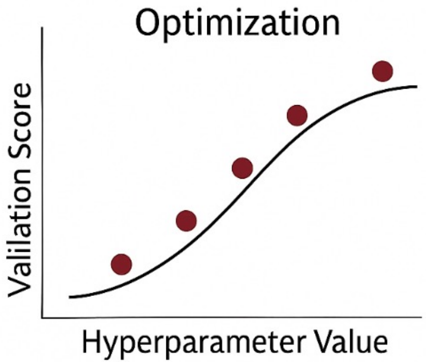
- Reliable performance measure

Modeling Concerns 2: *Hyperparameter* Trick: *Optuna is a Bayesian-based Tuning Tool*

OPTUNA

Optuna: A hyperparameter optimization framework

Cross-Validation	
Fold 1	Training Data
Fold 2	Validation Data
Fold 3	Validation Data
Fold 4	Validation Data



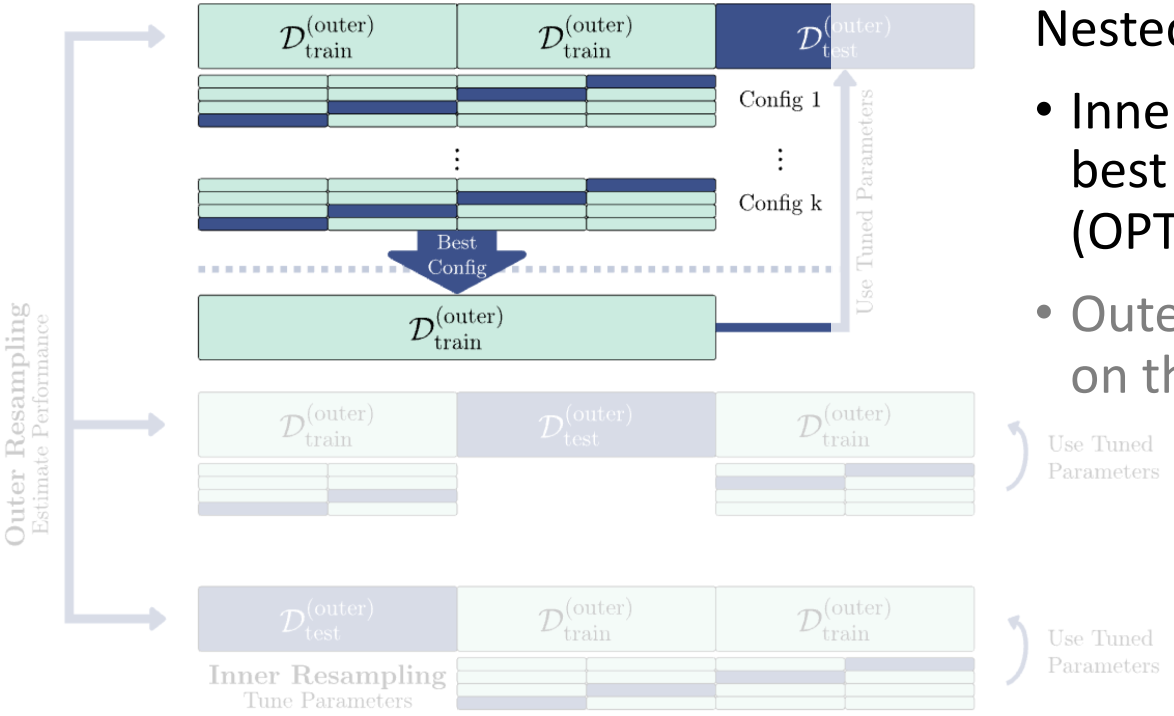
Bayesian optimization used to guide search

Optuna automatically searches on a set of hyperparameters, and trains/validates using k-fold cross-validation (CV)

Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. 2019. Optuna: A Next-generation Hyperparameter Optimization Framework. In KDD.

Modeling Concern 3: *Model Generalizability*
Trick: *Nested Resampling Scheme*

USC Mann
Alfred E. Mann School of Pharmacy and Pharmaceutical Sciences

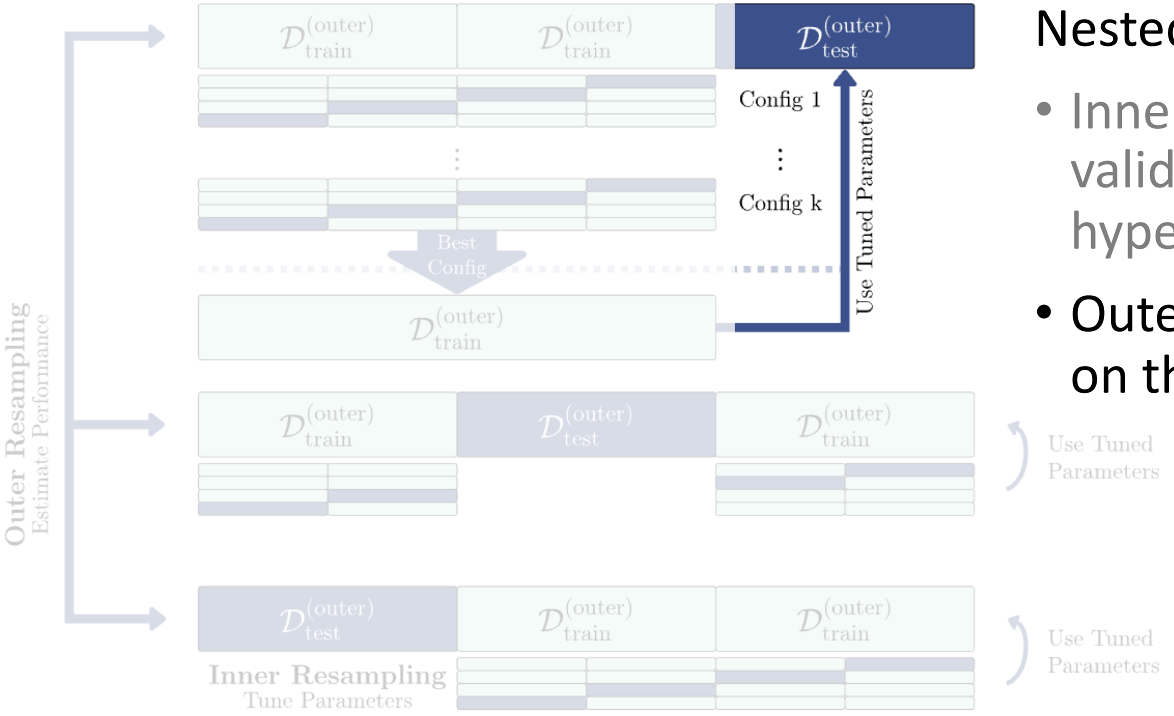


Nested: Inner + Outer

- Inner: A k-fold CV to find the best hyperparameter (OPTUNA)
- Outer: Measure performance on the test split

USC Mann
Alfred E. Mann School of Pharmacy and Pharmaceutical Sciences

Modeling Concern 3: *Model Generalizability*
Trick: *Nested Resampling Scheme*

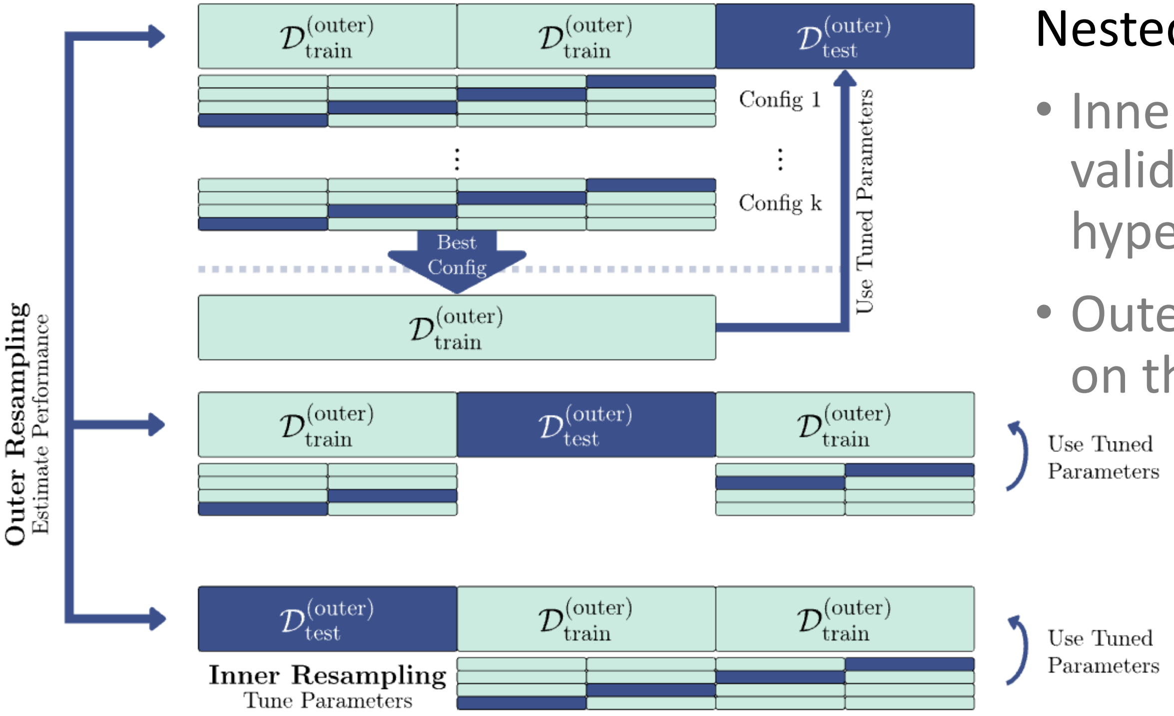


Nested: Inner + Outer

- Inner: A k-fold cross-validation to find the best hyperparameter (Optuna)
- Outer: Measure performance on the test split

USC Mann
Alfred E. Mann School of Pharmacy and Pharmaceutical Sciences

Modeling Concern 3: *Model Generalizability*
Trick: *Nested Resampling Scheme*



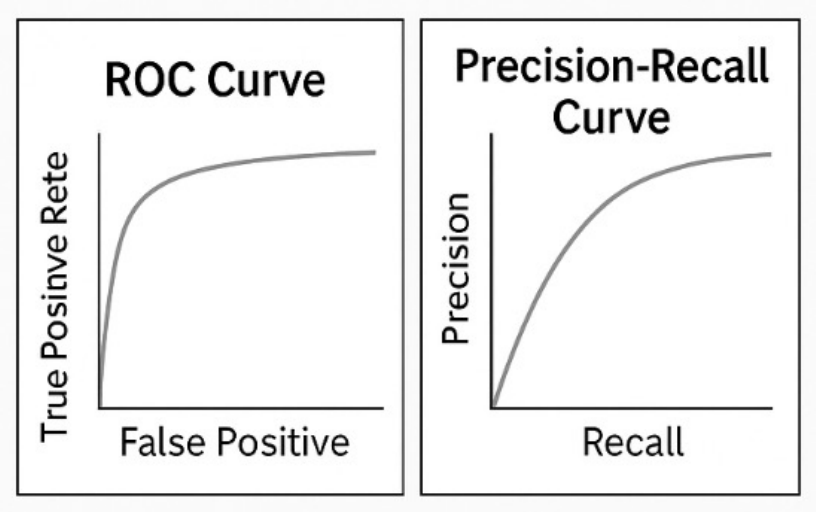
Nested: Inner + Outer

- Inner: A k-fold cross-validation to find the best hyperparameter
- Outer: Measure performance on the test split

Repeated Multiple Times

Performance Metrics Pool for Boostrapping

Modeling Concern 4: *Model Performance*
Trick: *ROC-AUC & PR-AUC*



ROC-AUC: Discrimination ability

- Insensitive to imbalanced data

PR-AUC: Ability in detecting minority class (positives) in imbalanced data

- Sensitive to imbalanced data

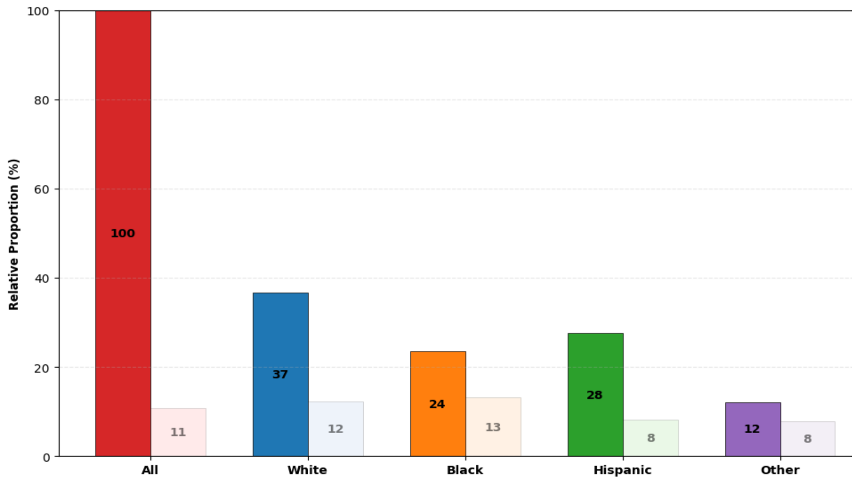
Outline

USC Mann
Alfred E. Mann School of Pharmacy and Pharmaceutical Sciences

- Background for Machine Learning and Racial Fairness Question
- Data Sources and Model Building Tricks
- **Performance Comparisons**
- Lessons from Practice

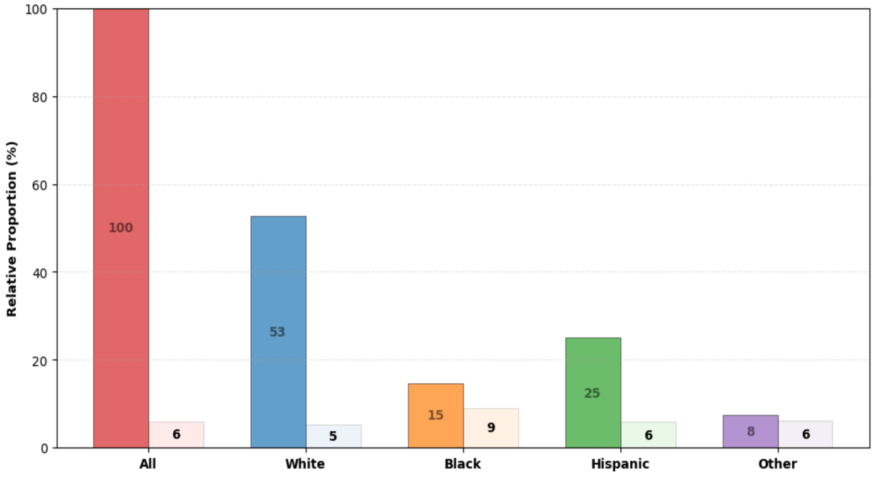
USCMann
Alfred E. Mann School of Pharmacy and Pharmaceutical Sciences

Glance at Imbalanced Data: Relative Proportion



CVD Prediction

- More White and Hispanic
- Black had highest prevalence

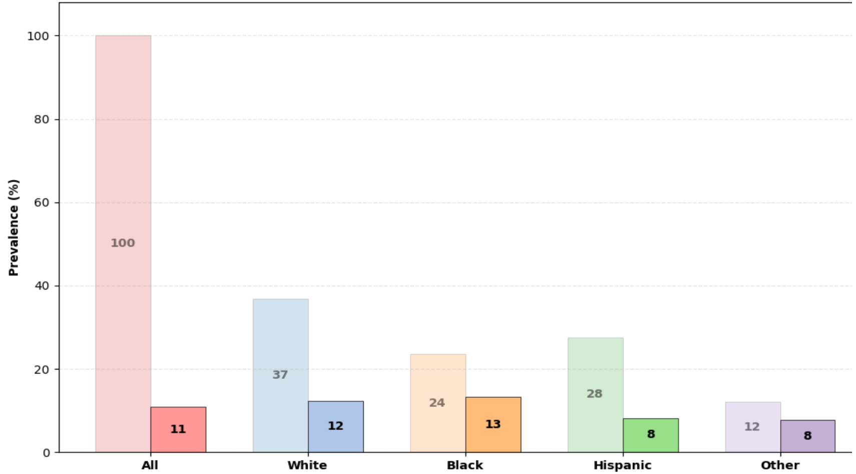


APO Prediction

- More White and Hispanic
- Black had highest prevalence

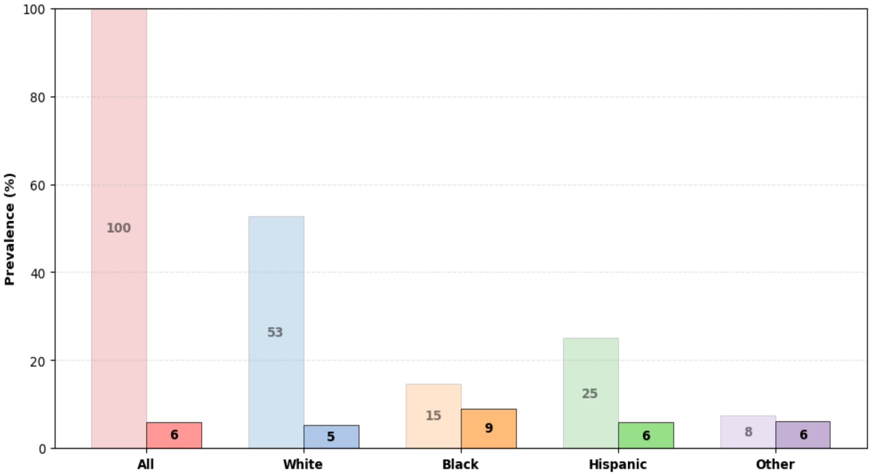
USC Mann
Alfred E. Mann School of Pharmacy and Pharmaceutical Sciences

Glance at Imbalanced Data: Prevalence of Outcome



CVD Prediction

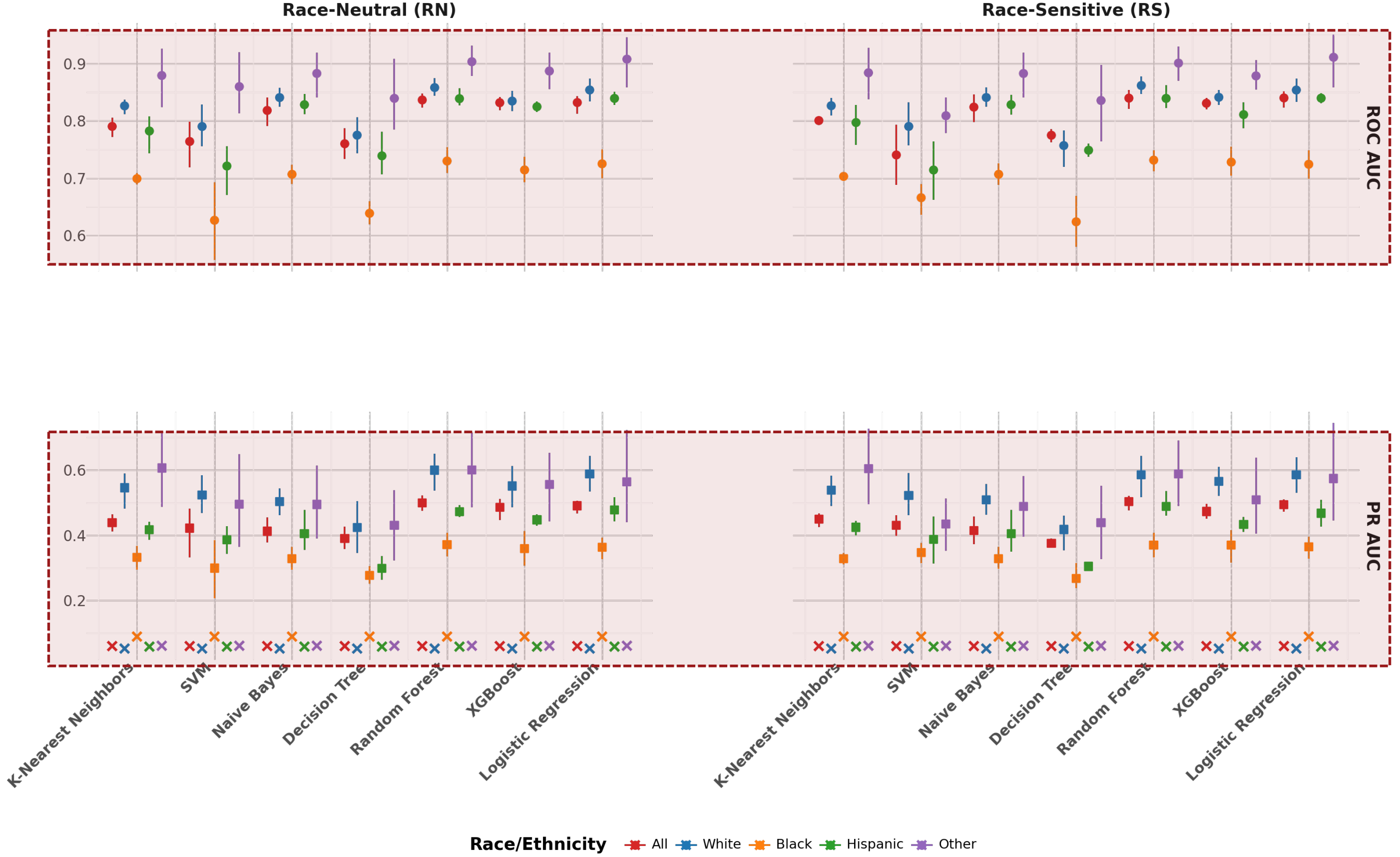
- More White and Hispanic
- Black had highest prevalence



APO Prediction

- More White and Hispanic
- Black had highest prevalence

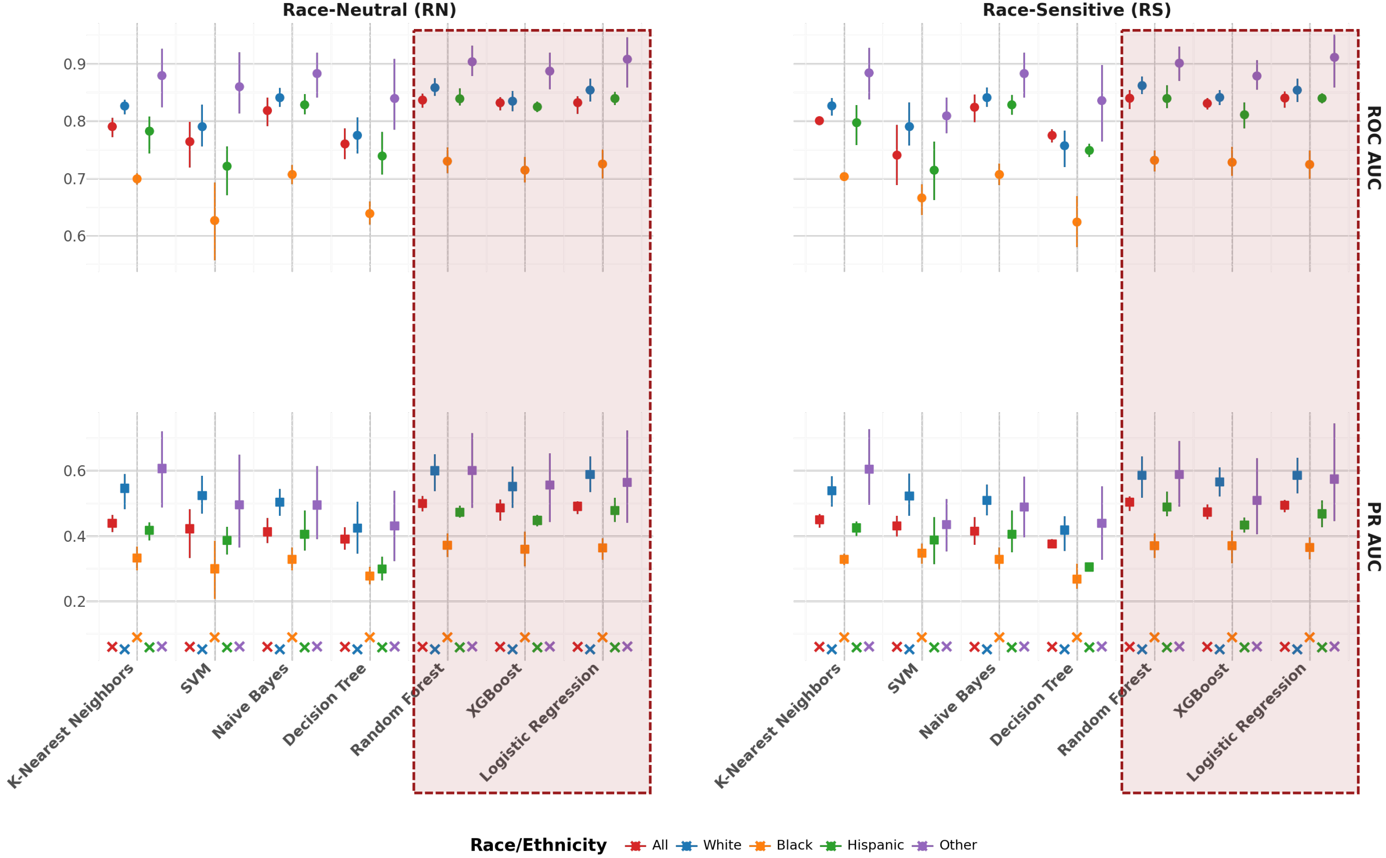
CVD Prediction: Race Matters? No



No Difference

- Tree-based learners performed better
- All-population model was modest with some tradeoffs
- Worst performance for black-specific models, though largest prevalence

CVD Prediction: Race Matters? No

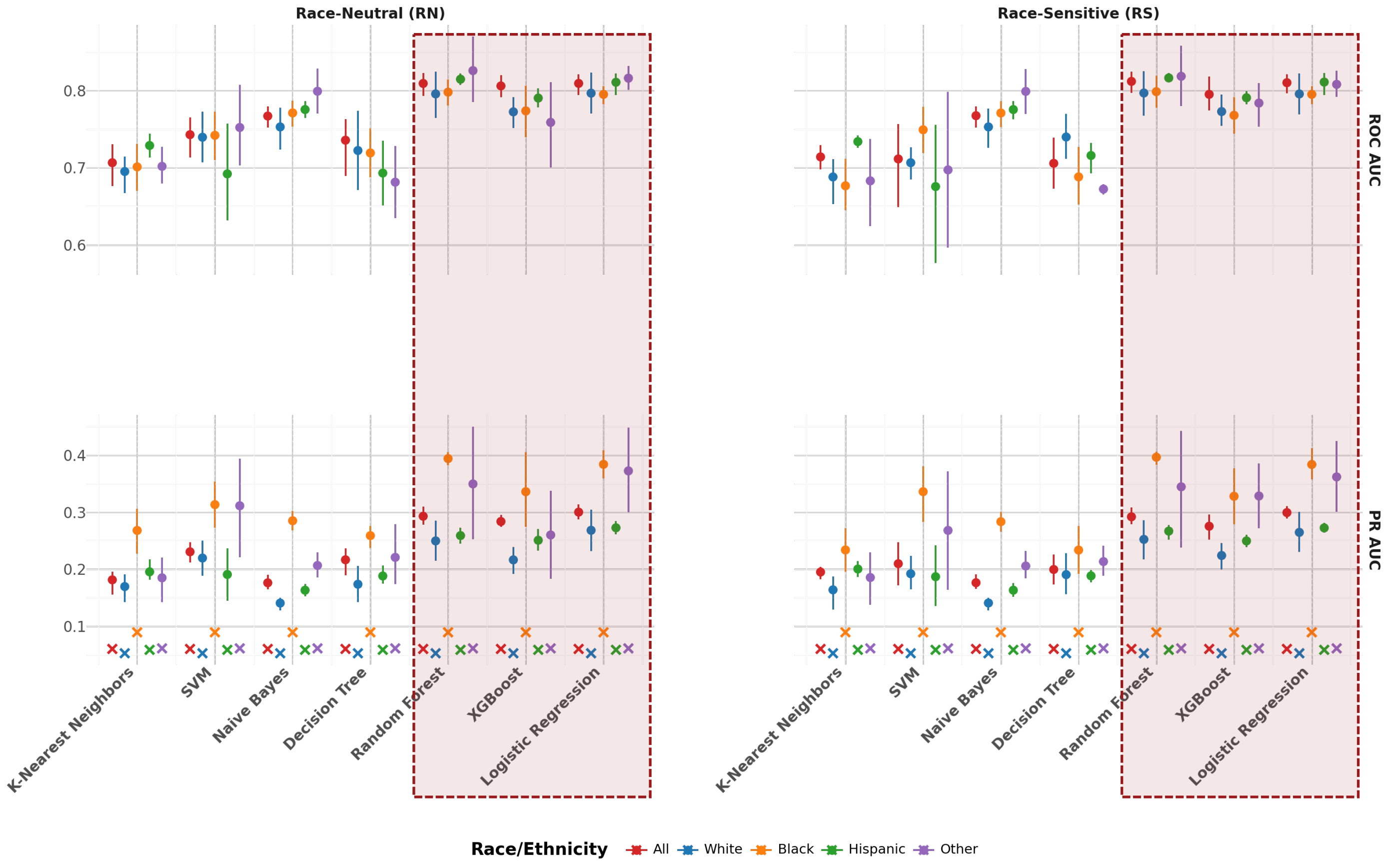


No Difference

- Tree-based learners performed better, similar to logistic reg.
- All-population model was modest with some tradeoffs
- Worst performance for black-specific models, though largest prevalence

USC Mann
Alfred E. Mann School of Pharmacy
and Pharmaceutical Sciences

APO Prediction: Race Matters? No



No Difference

- Worst performance for white-specific models

Outline

USC Mann
Alfred E. Mann School of Pharmacy and Pharmaceutical Sciences

- Background for Machine Learning and Racial Fairness Question
- Data Sources and Model Building Tricks
- Performance Comparisons
- Lessons from Practice

USC Mann
Alfred E. Mann School of Pharmacy and Pharmaceutical Sciences

Advanced models are not always better

JAMA Network Open

Original Investigation | Cardiology

Comparison of Machine Learning Methods With Traditional Models for Use of Administrative Claims With Electronic Medical Records to Predict Heart Failure Outcomes

Rishi J. Desai, MS, PhD; Shirley V. Wang, PhD; Muthiah Vaduganathan, MD, MPH; Thomas Evers, PhD; Sebastian Schneeweiss, MD, ScD

Machine learning methods offered only limited improvement over traditional logistic regression in predicting key HF outcomes.

- They may not be dominantly better than traditional models, e.g., Logistic Reg.
  - Cautiously deploying models in your daily work!

USC Mann
Alfred E. Mann School of Pharmacy and Pharmaceutical Sciences

Algorithms uncover unrecognized bias

PNAS RESEARCH ARTICLE | PSYCHOLOGICAL AND COGNITIVE SCIENCES OPEN ACCESS

People see more of their biases in algorithms

Begum Celiktutan[a], Romain Cadario[a], and Carey K. Morewedge[b,1]

Edited by Elke Weber, Princeton University, Princeton, NJ; received October 10, 2023; accepted March 19, 2024

April 10, 2024 | 121 (16) e2317602121 | https://doi.org/10.1073/pnas.2317602121

Significance

Algorithms incorporate biases in the human decisions that comprise their training data, which can amplify and codify discrimination. We examine whether algorithmic biases can be used to reveal and help correct undetected biases of the human decision-makers on which algorithms are trained. We show that people see more of their biases in the decisions of algorithms than in their own decisions. Because algorithms reveal more of their biases, people are also more likely to correct their biases when decisions are attributed to an algorithm than to themselves. Recognizing bias is a crucial first step for people and organizations motivated to reduce their biases. Our findings illustrate how to use algorithms as mirrors to reveal and debias human decision-making.

- *People are not aware of their biases when making decisions*
- *Their self-trained algorithms reveal more biases than their own decisions, as much as they see in other people*
- *People are more likely to correct their biases in algorithms*

USC Mann
Alfred E. Mann School of Pharmacy and Pharmaceutical Sciences

Guidance is helpful

npj | digital medicine Article
Published in partnership with Seoul National University Bundang Hospital
https://doi.org/10.1038/s41746-024-01245-y

Guidance for unbiased predictive information for healthcare decision-making and equity (GUIDE): considerations when race may be a prognostic factor

Check for updates

Keren Ladin[1,2], John Cuddeback[3], O. Kenrik Duru[4], Sharad Goel[5], William Harvey[6], Jinny G. Park[7], Jessica K. Paulus[8], Joyce Sackey[9], Richard Sharp[10], Ewout Steyerberg[11], Berk Ustun[12], David van Klaveren[7,13], Saul N. Weingart[6] & David M. Kent[7,14]

Prediction and Causal Inference are distinct

- Decision contexts matter
- Antidiscrimination principles matter
- Testing and identifying trade-offs in achieving equity-related goals matter

USC Mann
Alfred E. Mann School of Pharmacy and Pharmaceutical Sciences

Takeaways

Should we include race in the machine learning models?

IT DEPENDS ON CONTEXTS

- Evidence is mixed
- A guideline for model-building practice is needed

STILL RECOMMEND TESTING WITH CAUTION

- Uncover unrecognized bias in the data
- Understand the role of race in the question
- Carefully interpret the results