

Participatory Systems let users decide which data is used for predictions and can improve fairness, privacy & performance.



Paper



Repo

Participatory Systems for Personalized Prediction

Hailey Joren
Chirag Nagpal

UC San Diego
Carnegie Mellon University

Katherine Heller
Berk Ustun

Google
UC San Diego

Predictive systems use personal information to improve predictions

Age	<65 0	65-74 +1	≥75 +2
Sex	Female +1	Male 0	
CHF history	No 0	Yes +1	
Hypertension history	No 0	Yes +1	
Stroke/TIA/thromboembolism history	No 0	Yes +2	
Vascular disease history (prior MI, peripheral artery disease, or aortic plaque)	No 0	Yes +1	
Diabetes history	No 0	Yes +1	

- The model might require info like sex and age to make a prediction (e.g., **CHA₂DS₂-VASc Score for Stroke Risk**)
- These are included in model training to provide better predictions across heterogeneous groups

Using a predictive system that requires personal information may not improve performance and can lead to *worse* predictions

Prediction Performance for Sleep Apnea

Group Attributes		n	Error Personalized	Error Generic	Gain
age	sex	1156	19.4%	20.4%	+1.0%
age < 30	female	48	26.8%	38.1%	+11.3%
age < 30	male	49	26.7%	23.9%	-2.8%
age 30-60	female	304	29.1%	30.3%	+1.2%
age 30-60	male	447	15.2%	15.4%	+0.2%
age > 60	female	123	21.9%	19.3%	-2.6%
age > 60	male	181	8.2%	11.0%	+2.8%

Why is this bad?

- The personal info can be costly or sensitive – providing it should always improve performance.
- In some cases, users are getting a *worse* prediction because they provided personal info

We apply participatory systems to 6 real-world healthcare prediction problems and find benefits for fairness, privacy, and performance

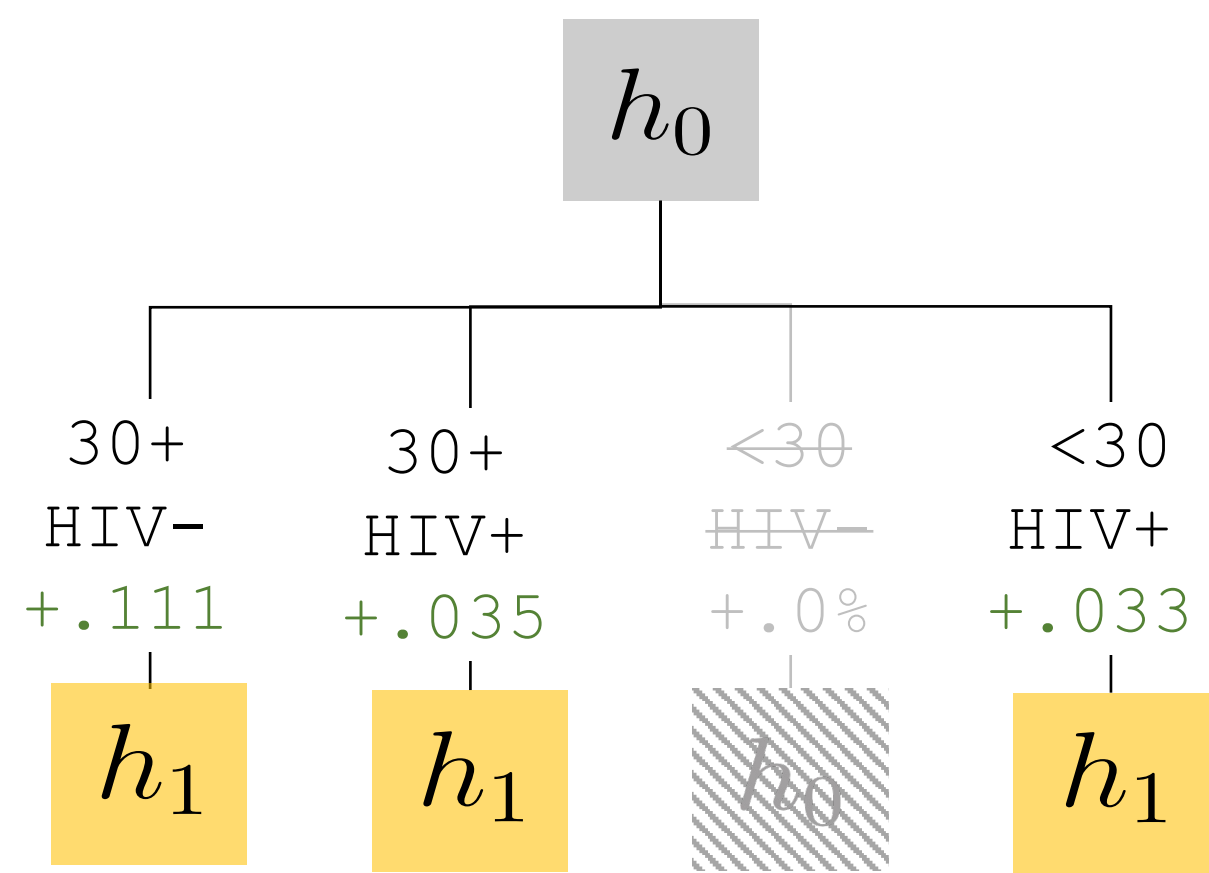
Case Study: ICU Mortality Prediction

	Logistic Regression	Minimal	Flat	Sequential
Test AUC	0.877	0.875	0.960	0.960
Groups Harmed	2	0	0	0
Max Gain / Group	0.019	0.015	0.139	0.139

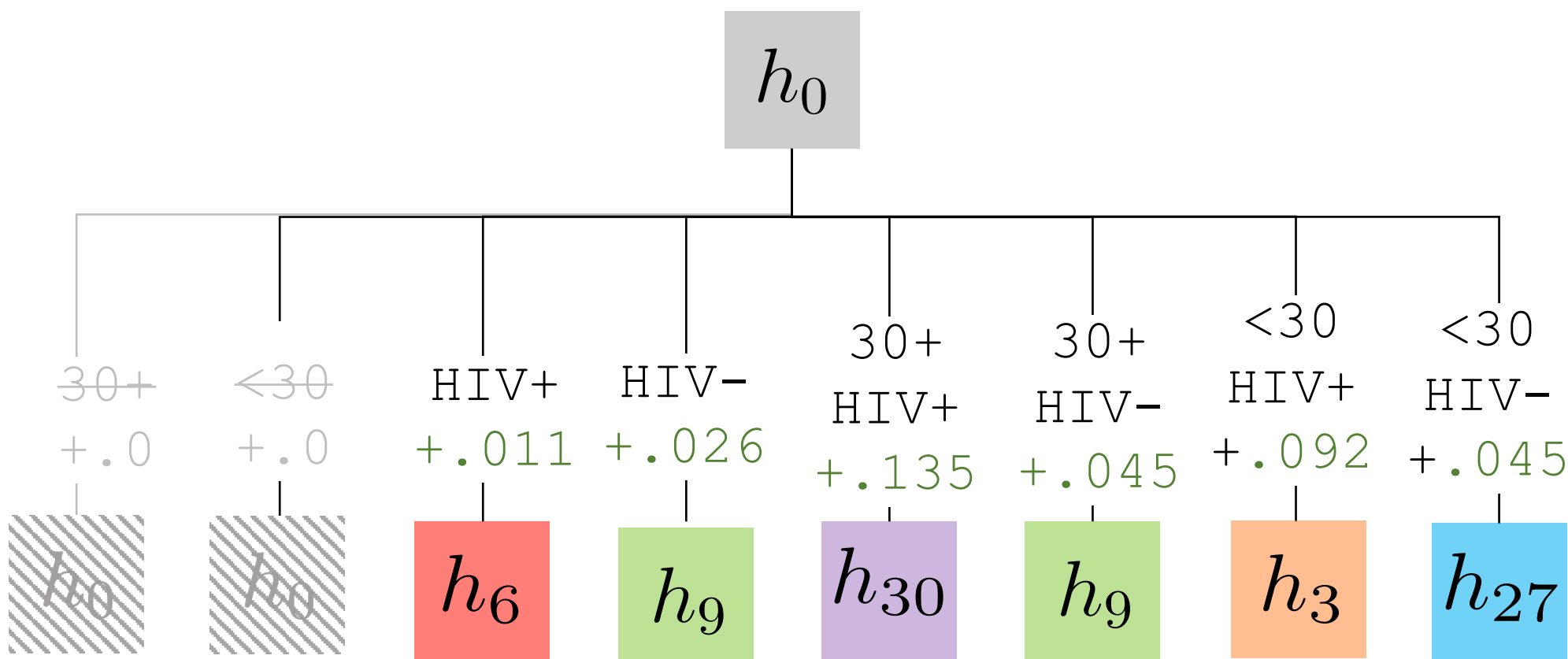
- Participatory systems achieve the best overall and group-level performance on 6/6 datasets
- The number of groups harmed by the original system is reduced in participatory systems
- The participatory systems do not solicit data when the performance gain does not pass a hypothesis test

We present a model-agnostic approach that uses multiple models to let users decide when to provide personal information

Minimal



Flat



Sequential

