Advancing Health Equity with Machine Learning

Vishwali Mhasawade

Motivation: Health Equity

"Minimizing avoidable disparities in health and its determinants–including but not limited to health care–between groups of people who have different levels of underlying social advantage or privilege, i.e., different levels of power, wealth, or prestige due to their positions in society relative to other groups." (Braveman, 2006)

Introduction

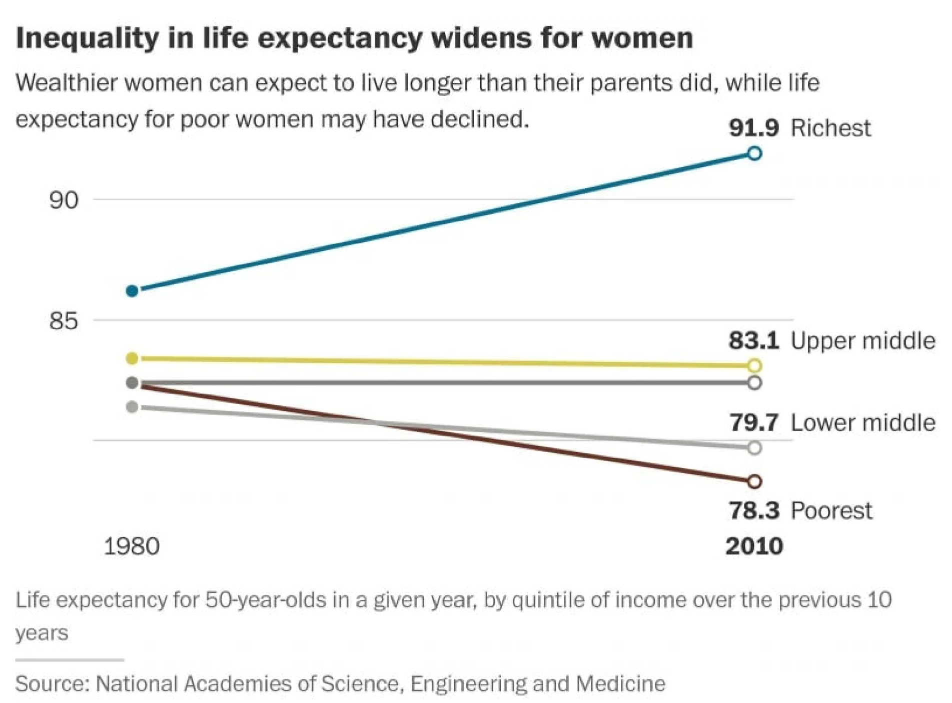

Figure 1:Growing inequalities for women in the United States (National Academies, 2015).

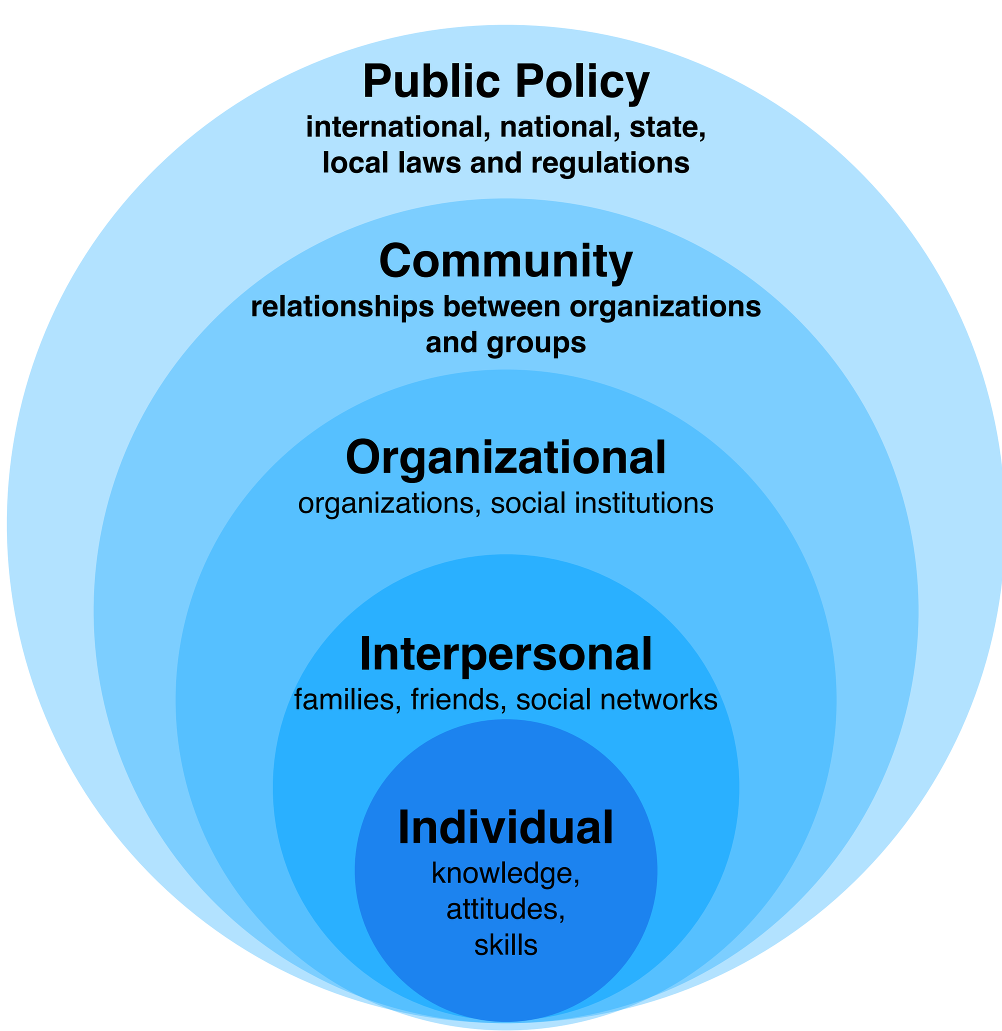

Figure 2:**The socioecological model of health** (Bronfenbrenner, 1977). Macro-properties (those above the individual level) are also key to understanding inequities and designing interventions that can reduce health inequities (Mhasawade et al., 2021)

Improving subgroup prediction by combining multiple sources of data (Mhasawade et al., 2020)

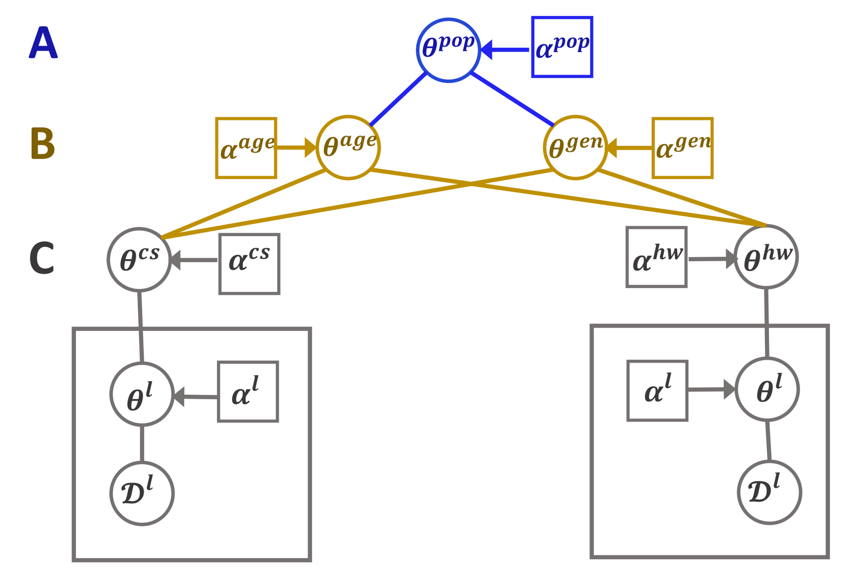

Figure 3:Population-aware hierarchical model; θ parameters at different nodes, $\mathcal{D}$ different data sets, α the priors. (A): Root level that represents invariant information across all data, (B): population parameters and information invariant to population-attributes (*age*) and (*gender*), (C): data set and environment-specific parameters and information (*cs* for citizen science and *hw* for healthworker facilitated datasets).

Table 1:AUC scores across population subgroups for Hutterite dataset.

	Age 5-15		Age 16-44		Age 45-64	
	Males	Females	Males	Females	Males	Females
TR	0.780	0.995	0.358	0.860	0.741	0.971
Hier	**0.957**	**1.000**	**0.576**	**1.000**	**0.890**	**0.971**

Table 2:AUC scores across population subgroups for Fluwatch dataset.

	Age 5-15		Age 16-44		Age 45-64	
	Males	Females	Males	Females	Males	Females
TR	0.762	0.726	0.808	0.708	0.293	0.678
Hier	**0.747**	**0.928**	**0.787**	**0.757**	**0.757**	**0.767**

Main contribution

Improving prediction by subgroup through combining multiple sources of data and reducing disparities in prediction by incorporating social determinants of health.

Incorporating social determinants for mitigating disparate model predictions (Mhasawade and Chunara, 2021)

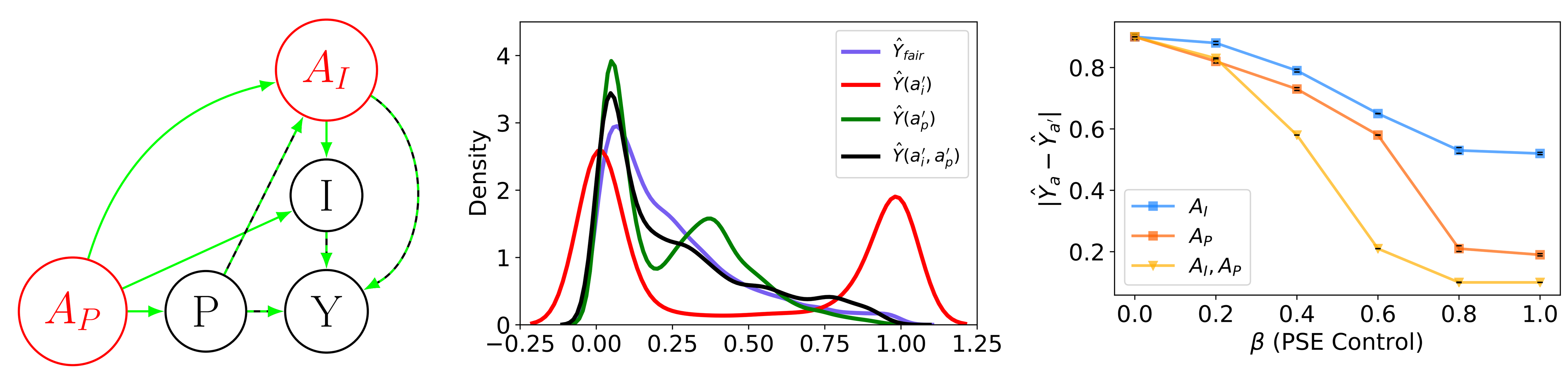

Figure 4:*Left*: Causal graph with multi-level sensitive attributes A_I, A_P. Macro-level variables, A_P (e.g. neighborhood SES), P (e.g. other zipcode level factors), and individual-level ones, A_I (e.g. perceived racial discrimination), I, affect the outcome Y (e.g. a health behavior), unfair paths are represented in green, *Center:* Density of $\hat{Y}$, *Right:* path-specific unfairness, $|\hat{Y}_a - \hat{Y}_{a'}|$ controlling for the effects of just A_I (blue), A_P (orange) and both A_I, A_P (yellow).

Conclusion

- Practitioners can save effort and cost by only labeling a proportion of data and combining data with other datasets to improve prediction.
- Using data from multiple sources for improving predictions by subgroup, thus, reducing disparities in model predictions.
- Our work extends algorithmic fairness to account for the multi-level and socially-constructed nature of forces that shape unfairness.
- We illustrate the importance of accounting for macro-level sensitive attributes by exhibiting residual unfairness if they are not accounted for.

References

Paula Braveman. Health disparities and health equity: concepts and measurement. *Annu. Rev. Public Health*, 27:167–194, 2006.

Urie Bronfenbrenner. Toward an experimental ecology of human development. *American psychologist*, 32(7):513, 1977.

Vishwali Mhasawade and Rumi Chunara. Causal multi-level fairness. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, pages 784–794, 2021.

Vishwali Mhasawade, Nabeel Abdur Rehman, and Rumi Chunara. Population-aware hierarchical bayesian domain adaptation via multi-component invariant learning. In *Proceedings of the ACM Conference on Health, Inference, and Learning*, pages 182–192, 2020.

Vishwali Mhasawade, Yuan Zhao, and Rumi Chunara. Machine learning and algorithmic fairness in public and population health. *Nature Machine Intelligence*, 3(8):659–666, 2021.

National Academies. *The growing gap in life expectancy by income: Implications for federal programs and policy responses.* National Academies Press, 2015.

NEW YORK UNIVERSITY

THE UNIVERSITY OF CHICAGO DATA SCIENCE INSTITUTE

Rising Stars in Data Science

SCAN ME