Bias Prevention in Global Adoption of Artificial Intelligence Based Health Management Systems

Aditya Sood, Ayush Sangari
Emory University School of Medicine

Introduction: Artificial intelligence is now capable of being used as a screening tool for a variety of purposes: distinguishing between benign nevi and melanoma, assessing diabetic and pressure ulcers, improving psoriasis and other inflammatory skin disease classifications, and more. Dermatology's embrace of artificial intelligence is good for both practitioners, as well as for patients. Our clinics have a mean wait time of 56 days. Accurate screening tools will inevitably allow for us to be able to better manage our patients, as well as see more of them. This technology offers massive benefits globally, but narrow artificial intelligence models pose massive bias when used on patients dissimilar to those which the datasets were created.

Objectives: This study analyses current medical policy practices with regards to the bias prevention in artificial intelligence systems to mitigate unforeseen bias in global adoption of AI based screening technologies.

Methods: This study began by surveying existing regulations and guidelines for the use of artificial intelligence in clinical care from prominent regulatory agencies, non-profits, and think-tanks internationally. These existing regulations and guidelines were then analyzed based upon their abilities to address bias prevention when applying artificial intelligence systems in diverse patient populations. New recommendations were also proposed to meet identified gaps between existing frameworks and current global needs.

Results:
Best practices to avoid bias in global adoption of artificial intelligence based health systems involve a focus on two topics: dataset expansion and model explainability.
AI programs are only as good as the dataset they are trained on. For this reason, the training dataset size and variety matter. Those currently being used to train programs are often vastly over-represented by type I-II skin patients. One of the largest open source archives of skin images - the International Skin Imaging Collaboration: Melanoma Project – is predominantly composed of fair skinned patients.

AI systems based on homogenous datasets do poorly on non-matching data, leading to poor outcomes and increased mortality rates.

This study also concluded that an integral aspect of bias prevention in adopting artificial intelligence systems globally is ensuring that the results of artificial intelligence systems are explainable and interpretable, allowing them to catch instances where bias may exist. Current enforcement of AI explainability depends on either satisfying a set of mathematical constraints or satisfying a general yet vague set of criteria. This study proposes that such systems be able to answer a baseline set of five fundamental questions in an accessible manner.

1. How was the information presented to explain the model obtained?
2. How does each inputted attribute/feature contribute to the final prediction/output?
3. What are some instances that the model considers to be similar to a specific instance?

4. What are the defining characteristics of each identified output class?
5. How can a provided instance be modified so that the model output is a certain class?

This provides the opportunity for different explanatory techniques to satisfy the same question. Furthermore, these questions are uniquely framed in terms that are easily understood by patients and physicians alike.