When Algorithmic Predictions Use Human-Generated Data: A Bias-Aware Classification Algorithm for Breast Cancer Diagnosis

“Algorithms that learn from human decisions will also learn human mistakes, such as overtesting and overdiagnosis, failing to notice people who lack access to care, undertesting those who cannot pay, and mirroring race or gender biases. Ignoring these facts will result in automating and even magnifying problems in our current health system. Noticing and undoing these problems requires a deep familiarity with clinical decisions and the data they produce—a reality that highlights the importance of viewing algorithms as thinking partners, rather than replacements, for doctors.” —Obermeyer 2017, New England Journal of Medicine

Motivation: Major advancements in information technology (IT) in the last decade have produced an unprecedented rise in readily available and complex datasets. Such ubiquity of data has facilitated an increased use of algorithms—analytical tools that use available data and produce an output (such as a prediction)—by the decision makers in social and economic transactions. The pervasiveness of data promises significant opportunities to improve decision making in every conceivable domain with the help of the algorithmic approach and efforts are already underway in domains ranging from travel-pricing forecasts to the prediction of recidivism, and to diagnostic decisions in medicine.

An important advantage of the algorithmic approach to decision making is that it may not suffer from the cognitive limitations or biases that are typically associated with human decision making. However, when the input data used by the algorithms are generated by human beings, even algorithms become susceptible to human biases. The bias in input data would likely diminish the algorithm's performance. For a specific example in healthcare, consider an algorithm that predicts the probability of breast cancer for a patient based on the patient's clinical characteristics and the radiologist's assessment of the patient's mammogram. However, the radiologist often accesses the patient's clinical characteristics while providing her mammogram assessment, and her assessment is influenced by the patient's clinical characteristics. The phenomenon—the use of human-generated decisions as inputs for prediction algorithms—arises in many contexts including loan approvals, job-hiring decisions, law-enforcement-related decisions, etc.
Algorithms that ignore bias in inputs may provide predictions that suffer from limitations akin to those provided by human beings. Furthermore, an algorithm that ignores the biases in the input data could also exacerbate the errors stemming from those biases. Therefore, designing an optimal decision-support algorithm that accounts for biases in the input data is critical. Consistent with this theme, in this paper, we examine the design and value of an algorithm that accounts for bias in the input data. Specifically, we develop a linear classifier that accounts for the human-generated bias in the input data. We contextualize the problem in healthcare and examine the design of a bias-aware algorithm for use in a clinical decision support system (CDSS). The CDSS is a decision aid to be used by referring physicians in mammography-based breast cancer diagnosis. Given the increased interest in the development of automated risk calculators for diagnosing breast cancer in recent years, we quantify the impact of bias, and provide numerous insights about the design of clinical CDSS for breast cancer diagnosis.

**Research Question:** We develop a linear classification algorithm that uses two information sources—the radiologist's mammogram assessment and the clinical-risk information, where the mammogram assessment could be biased by the clinical-risk information—and determines whether a patient has cancer. We seek to answer three key research questions: (i) What is the optimal design of a linear classification algorithm in the presence of radiologist bias? (ii) How does the bias affect the algorithm's performance and design? (iii) Should the clinical-risk information be used at all in the diagnostic process given its potential to bias the radiologist? We answer the above questions by first developing and analyzing a theoretical model of the breast-cancer-diagnosis context. We then quantify the theoretical results using real-life data widely used by the breast cancer medical community.

**Findings:** We develop and show that a bias-aware algorithm can eliminate the adverse impact of
bias if the error in the mammogram assessment due to radiologist's bias has no variance. On the other hand, in the presence of error variance, the adverse impact of bias can be mitigated, but not eliminated, by the bias-aware algorithm. The bias-aware algorithm assigns less (more) weight to the clinical-risk information (radiologist's mammogram assessment) when the mean error increases (decreases), but the reverse happens when the error variance increases. Using point estimates obtained from mammography practice and the medical literature, we show that the bias-aware algorithm can significantly improve the expected patient life years or the accuracy of decisions based on mammography.

**Contributions:** First, our approach to designing a bias-aware classification algorithm and our analysis of the impact of bias on the algorithm and its performance are new to the literature on classification algorithms. We contribute to this literature by (i) explicitly modeling the source and mechanics of noise (introduced by human biases) and (ii) separating the noise into systematic and random parts. Consequently, we are able to provide deep theoretical insights into how the input bias affects the optimal bias-aware algorithm design and the algorithm's performance. Second, this research quantifies the value of bias-aware algorithms in the breast-cancer-diagnosis context. Rapid digitalization of healthcare and the growing interest in the role of cognitive biases in medical decision making make healthcare an ideal application domain for bias-aware algorithms. The medical literature recognizes that CDSS are useful in medical decision making, potentially also in mitigating potential harms of biases. Despite the interest in decision biases, to our knowledge, the literature has not rigorously examined the design of bias-aware algorithms and their value in mitigating the impact of bias. Our findings contribute to not only the design of a new class of breast cancer risk calculators that account for radiologists' bias, but also the value of such calculators in terms of decision accuracy and utility.