

ETHICAL REFLECTION.

Ensuring Fairness in Cancer Screening Predictions: Addressing Bias with IBM AI Fairness 360

Introduction

Predictive models, such as Random Forest classifiers, are increasingly used in healthcare to assist in cancer screening—particularly in distinguishing between benign and malignant tumors. While these models offer promising support for clinicians, they can introduce or amplify biases if the underlying data is flawed.

Potential Biases in Cancer Screening Datasets

1. Demographic Imbalance

If the dataset contains significantly more examples from certain population groups (e.g., middle-aged women from urban areas) and fewer from others (e.g., elderly patients, minority groups, or males), the model may generalize poorly for underrepresented populations. This could lead to higher false negatives or misdiagnoses for these groups.

2. Healthcare Access Bias

Patients who undergo regular screenings often have better healthcare access. If the dataset is skewed toward such individuals, the model may perform poorly on patients with limited or inconsistent access to healthcare—typically lower-income or rural populations.

3. Labeling Bias

Diagnosis labels (benign/malignant) may be influenced by past diagnostic errors or the availability of diagnostic tools in certain facilities, causing mislabeling and model confusion.

Mitigating Bias with IBM AI Fairness 360

IBM AI Fairness 360 (AIF360) is an open-source toolkit designed to identify and reduce bias in machine learning models. In cancer prediction, it can be applied as follows:

1. Fairness Metrics

AIF360 offers metrics like statistical parity difference, equal opportunity difference, and disparate impact to quantify whether certain demographic groups are disproportionately misclassified.

2. Pre-processing Techniques

AIF360 can re-balance datasets using techniques like reweighing or disparate impact remover, helping to ensure equal representation across sensitive attributes (e.g., gender, age, ethnicity).

3. In-Processing and Post-Processing Corrections

The toolkit provides algorithms that adjust model training or outcomes to enforce fairness constraints—helping ensure that predictions do not unfairly disadvantage specific groups even after the model is trained.

Conclusion

In cancer screening applications, fairness is not just an ethical concern—it can directly impact patient lives. Addressing dataset and algorithmic bias using tools like IBM AI Fairness 360 ensures that predictive models support equitable, reliable, and clinically responsible decision-making. As AI becomes more embedded in healthcare, such fairness interventions will be critical in building trust and delivering better outcomes for all.