Bias Mitigation Comparison Report

Study: Bias Mitigation Method Comparison

The study is designed to systematically evaluate bias mitigation methods implemented by the user. After bias amplification through inductive transfer learning, mitigation methods are applied on these biased models and assessed by their effectiveness under different levels of bias.

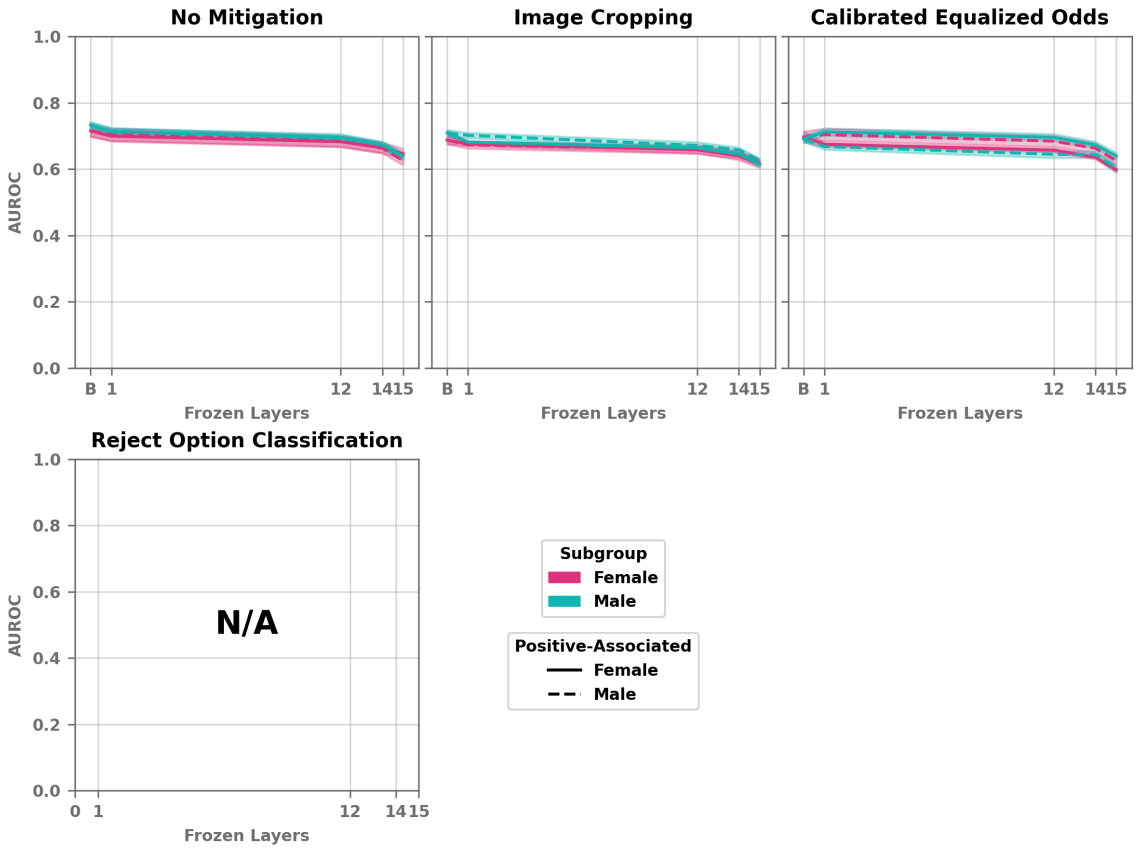
Bias Ampilification Approach: Inductive Transfer Learning

The inductive transfer learning approach applies a two-step transfer learning approach where the AI is trained to classify patient attributes during the first step. The AI is then fine-tuned to perform clinical tasks during the second step. Additional controls over the degree to which bias is amplified is taken by number of frozen layers during fine-tuning in second step.

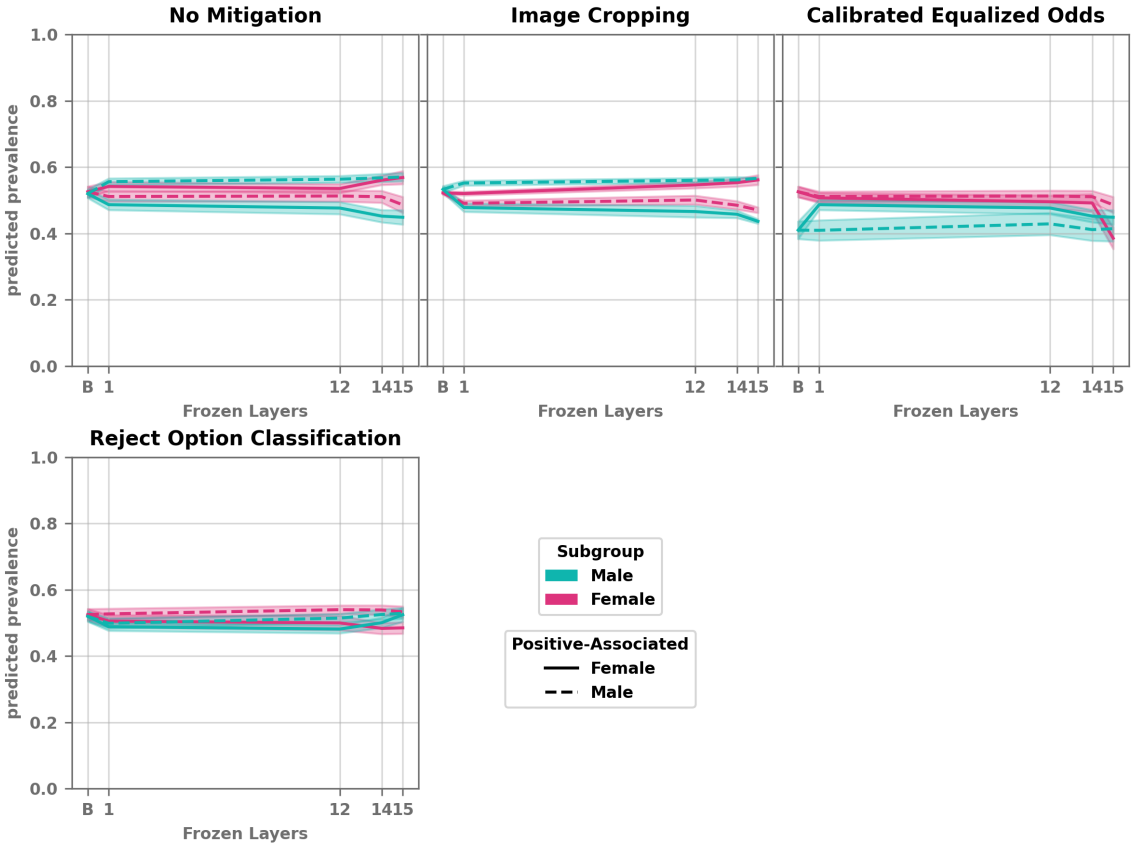
Results

Each figure under this section presents one metric result respectively. Inside each figure, the first subplot presents the amplified bias (without mitigation), while the rest subplots show results from different implemented mitigation methods. For these experiments, the positive-associated subgroup refers to the subgroup associated with the same model output during extra transfer learning step. The x-axis indicates the number of layers being frozen during the final model fine-tune step, while B indicates the baseline model.

1. AUROC



2. predicted prevalence



Report Metadata

Report generated by: bias.myti.report v1.0

Date: 2024-03-11

Time: 18:36:24