Y1: performance of single vs ensemble model on subgroup-by-race: non-white 120 one-model-for-all one-model-per-subgroup ensemble-on-all ensemble-on-subgroup 100 100 100 100 100 100 100 97.8 97.8 100 -91.3 80 80 80 66.7 66.7 Score(%) 60 · 50 40 33.3 20 0 precision f1 recall accuracy