COT: ACC 0.40 / AUROC 0.58 / ECE 0.55 95 - · 65 -60 -31 16 Confidence 55 - -45 -6 · correct answer wrong answer

Count