COT: ACC 0.40 / AUROC 0.81 / ECE 0.32 238<sup>l</sup> - -65 -60 -Confidence 55 -50 -45 - · correct answer wrong answer **Count**