

COT: ACC 0.67 / AUROC 0.61 / ECE 0.29

Count



wrong answer



correct answer

Confidence (%)

