

**COT: ACC 0.66 / AUROC 0.53 / ECE 0.23**

**Count**

**wrong answer**  
**correct answer**

**50 60 70 80 90 100**  
**Confidence (%)**

