COT: ACC 0.41 / AUROC 0.50 / ECE 0.46

Confidence (%) vs Count

- 10 / 9 (≈98%)
- 802 / 525 (≈93%)
- 763 / 636 (≈87%)
- 188 / 86 (≈82%)
- 79 / 36 (≈77%)
- 62 / 28 (≈72%)
- 43 / 28 (≈67%)
- 39 / 14 (≈62%)
- 26 / 17 (≈57%)
- 18 / 22 (≈52%)
- 4 / 8 (≈47%)
- 2 / 1 (≈42%)
- 2 (≈37%)

Legend:
- correct answer (blue)
- wrong answer (red)