Performance Metrics & Benchmarks

MedQA: 87% accuracy

PubMedQA: 78% F1

MMLU-Medical: 91%

AUROC, Sensitivity, Specificity