

Variance Stability Ranking of Evaluation Metrics (All Scenarios Including UC4)

Lower average rank indicates better variance stability across 6 scenarios

