Metrics comparison on bias balanced dataset

