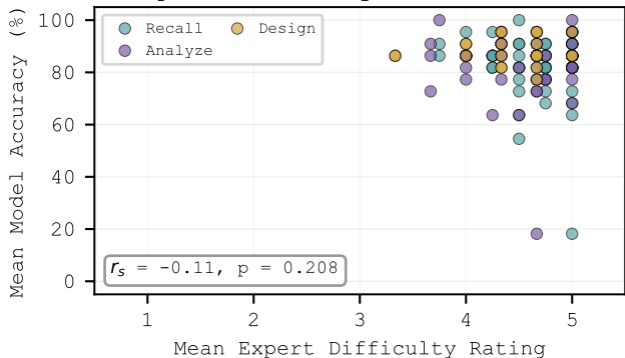


(A) Expert Difficulty vs. Model Accuracy



(B) Impact of Expert Quality Flags

