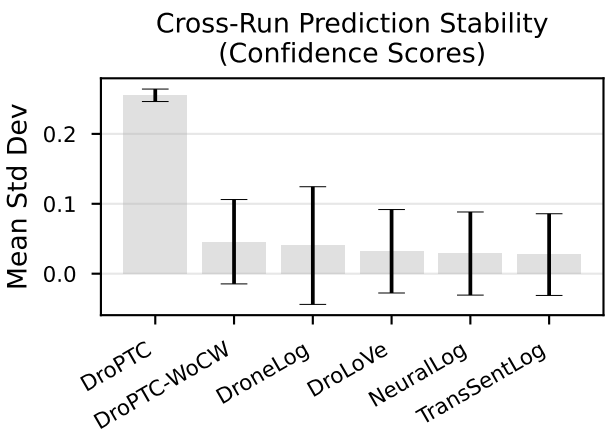
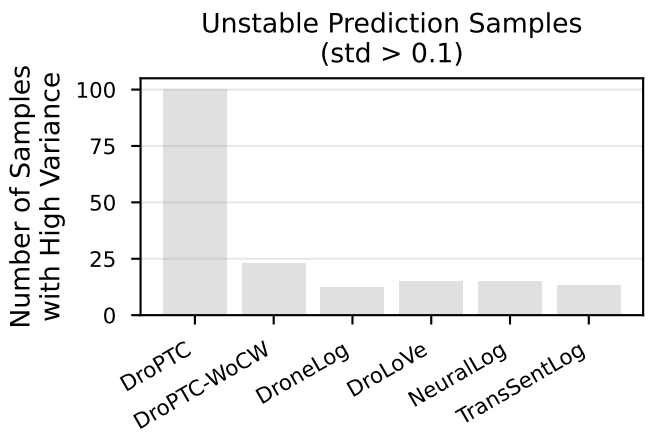
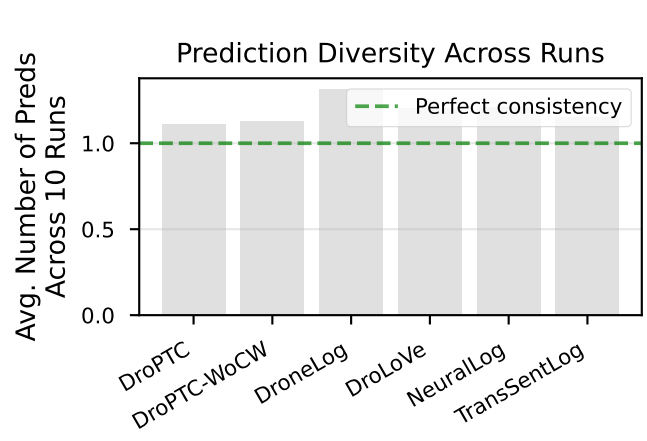

Cross-Run Prediction Stability
(Confidence Scores)
Mean Std Dev
0.2
0.1
0.0
DroPTC
DroPTC-WoCW
DroneLog
DroLoVe
NeuralLog
TransSentLog


Unstable Prediction Samples
(std > 0.1)
Number of Samples
with High Variance
100
75
50
25
0
DroPTC
DroPTC-WoCW
DroneLog
DroLoVe
NeuralLog
TransSentLog


Prediction Diversity Across Runs
Avg. Number of Preds
Across 10 Runs
Perfect consistency
1.0
0.5
0.0
DroPTC
DroPTC-WoCW
DroneLog
DroLoVe
NeuralLog
TransSentLog