Findings, N=2347 True Predicted 1.0 8.0 0.6 0.4 0.2 0.0 bleu4 rougeL bertscore f1radgraph f1chexbert