

## Appendix

Precision, recall, and  $F_1$  scores and the related confidence intervals for the four models discussed are provided below (BASELINE, BASELINE-OPT, SNORKEL (500), and UMLS-EDA (50)).







