

Dataset	Performance [%]	
	LSTM	RoBERTa
Anemia	88 ± 10.5	86 ± 9.2
Diabetes	81 ± 10.4	76 ± 4.2
IMDB	90 ± 5.2	95 ± 11.2
SNLI	78 ± 10.4	91 ± 7.4
SST	82 ± 4.3	94 ± 6.5
bAbI-1	100 ± 0	100 ± 0
bAbI-2	68 ± 9.5	100 ± 0
bAbI-3	60 ± 22.3	81 ± 24.3

Table 1: Average performance over 5 seeds, with standard deviation shown after \pm .