Dataset	Performance [%]	
	LSTM	RoBERTa
Anemia	$88^{+1.1}_{-2.2}$	$86^{+0.6}_{-0.7}$
Diabetes	$81^{+2.2}_{-2.9}$	$76^{+0.7}_{-0.6}$
IMDB	$90^{+0.4}_{-0.7}$	$95^{+0.2}_{-0.2}$
SNLI	$78^{+0.2}_{-0.3}$	$91^{+0.1}_{-0.1}$
SST	$82^{+0.6}_{-1.0}$	$94^{+0.3}_{-0.3}$
bAbI-1	$100^{+0.0}_{-0.0}$	$100^{+0.0}_{-0.0}$
bAbI-2	$68^{+9.1}_{-19.1}$	$100^{+0.1}_{-0.1}$
bAbI-3	$60^{+6.5}_{-4.9}$	$81^{+6.8}_{-20.0}$

Table 1: Average performance over 5 seeds, with a 95% confidence interval.