

<b>Metric</b>	<b>P1</b>	<b>P2</b>	<b>P3</b>	<b>P4</b>	<b>P5</b>	<b>P6</b>	<b>P7</b>	<b>P8</b>	<b>JUN</b>	<b>SEN</b>	<b>EXP</b>
Sensitivity	0.62	0.56	0.50	0.56	0.69	0.38	0.62	0.38	0.69	0.62	0.81
Specificity	0.81	0.59	0.69	0.45	0.25	0.67	0.46	0.85	0.73	0.71	0.73
F1 Score	0.24	0.12	0.13	0.09	0.08	0.09	0.10	0.17	0.19	0.17	0.23
Accuracy	0.80	0.59	0.68	0.46	0.27	0.65	0.46	0.82	0.72	0.70	0.74

Supplementary Table2: Performance metrics of 8 prompts and three researchers