

S3A Table. Estimates (with 95% confidence interval (CI)) for the Model Performance

Metrics in the test set.

Performance indicators [95% CI]/ Models	Multinomial LR	Multinomial LR with interaction terms	XGBoost	SVM	ANN
Accuracy	0.689 ⁴ [0.684, 0.695]	0.690 ³ [0.684, 0.696]	0.694 ² [0.688, 0.699]	0.688 ⁵ [0.683, 0.694]	0.698 ¹ [0.693, 0.704]
Cohen-Kappa coefficient	0.518 ³ [0.509, 0.526]	0.517 ⁴ [0.508, 0.526]	0.526 ¹ [0.517, 0.534]	0.516 ⁵ [0.507, 0.524]	0.526 ¹ [0.517, 0.535]
Matthews correlation coefficient	0.522 ⁴ [0.514, 0.531]	0.523 ³ [0.513, 0.531]	0.527 ¹ [0.519, 0.536]	0.522 ⁴ [0.514, 0.531]	0.527 ¹ [0.518, 0.536]
F1-score					
<i>mRS 0–2</i>	0.796 ² [0.790, 0.801]	0.795 ³ [0.789, 0.800]	0.794 ⁴ [0.788, 0.800]	0.797 ¹ [0.791, 0.803]	0.793 ⁵ [0.788, 0.799]
<i>mRS 3–5</i>	0.554 ³ [0.545, 0.563]	0.552 ⁴ [0.542, 0.562]	0.577 ² [0.568, 0.586]	0.541 ⁵ [0.532, 0.551]	0.603 ¹ [0.594, 0.611]
<i>mRS 6</i>	0.667 ³ [0.656, 0.676]	0.669 ² [0.658, 0.679]	0.671 ¹ [0.660, 0.680]	0.667 ³ [0.657, 0.677]	0.657 ⁵ [0.646, 0.668]
<i>Macro average</i>	0.672 ³ [0.666, 0.678]	0.672 ³ [0.666, 0.678]	0.681 ² [0.675, 0.687]	0.668 ⁵ [0.663, 0.674]	0.685 ¹ [0.678, 0.691]
<i>Weighted average</i>	0.681 ³ [0.675, 0.687]	0.680 ⁴ [0.674, 0.686]	0.690 ² [0.684, 0.695]	0.677 ⁵ [0.671, 0.683]	0.696 ¹ [0.690, 0.702]
Sensitivity (Recall)					
<i>mRS 0–2</i>	0.848 ³ [0.842, 0.855]	0.859 ² [0.853, 0.865]	0.826 ⁵ [0.818, 0.833]	0.865 ¹ [0.859, 0.872]	0.830 ⁴ [0.823, 0.837]
<i>mRS 3–5</i>	0.489 ³ [0.479, 0.500]	0.483 ⁴ [0.474, 0.494]	0.534 ² [0.524, 0.545]	0.464 ⁵ [0.455, 0.475]	0.598 ¹ [0.588, 0.609]
<i>mRS 6</i>	0.710 ² [0.697, 0.722]	0.701 ⁴ [0.688, 0.713]	0.702 ³ [0.689, 0.715]	0.713 ¹ [0.700, 0.725]	0.607 ⁵ [0.592, 0.620]
<i>Macro average</i>	0.682 ² [0.677, 0.688]	0.681 ³ [0.675, 0.687]	0.687 ¹ [0.681, 0.693]	0.681 ³ [0.675, 0.687]	0.678 ⁵ [0.672, 0.684]
<i>Weighted average</i>	0.689 ⁴ [0.684, 0.695]	0.690 ³ [0.684, 0.696]	0.694 ² [0.688, 0.699]	0.688 ⁵ [0.683, 0.694]	0.698 ¹ [0.693, 0.704]
Precision (PPV)					
<i>mRS 0–2</i>	0.749 ³ [0.742, 0.757]	0.739 ⁴ [0.732, 0.747]	0.765 ¹ [0.758, 0.773]	0.739 ⁴ [0.732, 0.747]	0.760 ² [0.753, 0.768]

<i>mRS 3–5</i>	0.639 ³ [0.627, 0.650]	0.644 ² [0.631, 0.655]	0.628 ⁴ [0.618, 0.639]	0.650 ¹ [0.638, 0.661]	0.608 ⁵ [0.597, 0.618]
<i>mRS 6</i>	0.629 ⁴ [0.616, 0.641]	0.639 ³ [0.626, 0.651]	0.641 ² [0.629, 0.654]	0.626 ⁵ [0.613, 0.638]	0.717 ¹ [0.705, 0.730]
<i>Macro average</i>	0.672 ⁴ [0.666, 0.678]	0.674 ³ [0.667, 0.680]	0.678 ² [0.672, 0.684]	0.672 ⁴ [0.666, 0.678]	0.695 ¹ [0.689, 0.701]
<i>Weighted average</i>	0.683 ³ [0.678, 0.689]	0.683 ³ [0.677, 0.689]	0.689 ² [0.683, 0.695]	0.682 ⁵ [0.677, 0.689]	0.696 ¹ [0.690, 0.702]

^{1,2,3,4,5} are the ranking of the model prediction performances based on the respective metrics. PPV; positive predictive value.

S3B Table. Estimates (with 95% confidence interval (CI)) for the Model Performance

Metrics in the training set.

Performance indicators [95% CI]/ Models	Multinomial LR	Multinomial LR with interaction terms	XGBoost	SVM	ANN
Accuracy	0.687 ⁵ [0.683, 0.690]	0.690 ⁴ [0.686, 0.693]	0.704 ¹ [0.701, 0.707]	0.693 ³ [0.690, 0.696]	0.702 ² [0.699, 0.705]
Cohen-Kappa coefficient	0.514 ⁵ [0.508, 0.519]	0.517 ⁴ [0.512, 0.522]	0.541 ¹ [0.536, 0.546]	0.522 ³ [0.516, 0.526]	0.530 ² [0.525, 0.535]
Matthews correlation coefficient	0.517 ⁵ [0.512, 0.523]	0.522 ⁴ [0.517, 0.527]	0.543 ¹ [0.538, 0.548]	0.529 ³ [0.524, 0.534]	0.533 ² [0.528, 0.537]
F1-score					
<i>mRS 0–2</i>	0.792 ⁵ [0.788, 0.795]	0.793 ⁴ [0.790, 0.797]	0.796 ² [0.793, 0.800]	0.796 ² [0.792, 0.799]	0.799 ¹ [0.795, 0.802]
<i>mRS 3–5</i>	0.553 ⁴ [0.548, 0.559]	0.554 ³ [0.549, 0.560]	0.592 ² [0.587, 0.597]	0.551 ⁵ [0.546, 0.556]	0.605 ¹ [0.600, 0.610]
<i>mRS 6</i>	0.665 ⁴ [0.660, 0.671]	0.670 ³ [0.664, 0.675]	0.689 ¹ [0.684, 0.695]	0.676 ² [0.670, 0.681]	0.650 ⁵ [0.644, 0.656]
<i>Macro average</i>	0.670 ⁵ [0.667, 0.674]	0.672 ⁴ [0.669, 0.676]	0.693 ¹ [0.689, 0.696]	0.674 ³ [0.671, 0.678]	0.685 ² [0.681, 0.688]
<i>Weighted average</i>	0.679 ⁵ [0.675, 0.682]	0.681 ⁴ [0.677, 0.684]	0.700 ¹ [0.696, 0.703]	0.682 ³ [0.678, 0.685]	0.697 ² [0.694, 0.700]
Sensitivity (Recall)					
<i>mRS 0–2</i>	0.841 ⁴ [0.837, 0.845]	0.855 ³ [0.851, 0.859]	0.825 ⁵ [0.821, 0.828]	0.874 ¹ [0.870, 0.877]	0.861 ² [0.857, 0.864]
<i>mRS 3–5</i>	0.492 ³ [0.486, 0.498]	0.488 ⁴ [0.482, 0.494]	0.549 ² [0.543, 0.555]	0.474 ⁵ [0.468, 0.480]	0.590 ¹ [0.584, 0.596]
<i>mRS 6</i>	0.707 ² [0.699, 0.714]	0.700 ⁴ [0.693, 0.707]	0.724 ¹ [0.718, 0.731]	0.703 ³ [0.696, 0.711]	0.579 ⁵ [0.571, 0.587]
<i>Macro average</i>	0.680 ⁴ [0.676, 0.683]	0.681 ³ [0.678, 0.684]	0.699 ¹ [0.696, 0.703]	0.684 ² [0.680, 0.687]	0.676 ⁵ [0.673, 0.680]
<i>Weighted average</i>	0.687 ⁵ [0.683, 0.690]	0.690 ⁴ [0.686, 0.693]	0.704 ¹ [0.701, 0.707]	0.693 ³ [0.690, 0.696]	0.702 ² [0.699, 0.705]
Precision (PPV)					
<i>mRS 0–2</i>	0.748 ² [0.744, 0.753]	0.740 ⁴ [0.735, 0.745]	0.770 ¹ [0.766, 0.775]	0.730 ⁵ [0.726, 0.735]	0.745 ³ [0.741, 0.750]

<i>mRS 3–5</i>	0.632 ⁴ [0.626, 0.639]	0.641 ³ [0.634, 0.647]	0.642 ² [0.636, 0.648]	0.659 ¹ [0.652, 0.666]	0.621 ⁵ [0.616, 0.627]
<i>mRS 6</i>	0.628 ⁵ [0.621, 0.635]	0.642 ⁴ [0.635, 0.649]	0.657 ² [0.650, 0.664]	0.650 ³ [0.643, 0.657]	0.741 ¹ [0.734, 0.748]
<i>Macro average</i>	0.669 ⁵ [0.666, 0.673]	0.674 ⁴ [0.671, 0.678]	0.690 ² [0.687, 0.693]	0.680 ³ [0.676, 0.683]	0.703 ¹ [0.699, 0.706]
<i>Weighted average</i>	0.681 ⁵ [0.677, 0.684]	0.683 ⁴ [0.680, 0.686]	0.700 ¹ [0.696, 0.703]	0.687 ³ [0.684, 0.691]	0.700 ¹ [0.696, 0.703]

^{1,2,3,4,5} are the ranking of the model prediction performances based on the respective metrics. PPV; positive predictive value.