

**Multimedia Appendix 3.** General department guidance accuracy. This figure presents the accuracy rates across three major hospital department categories. The vertical axis represents accuracy in percentage (%). Each bar within the subplots (a), (b), and (c) corresponds to accuracy values for individual runs (Run1, Run2, Run3) within each department category, respectively. Comparisons between runs within each department category are conducted using the Wilcoxon test, with p-values shown above the bars to indicate statistical significance (ns for non-significant results). The black dashed line represents the overall average accuracy across all departments.