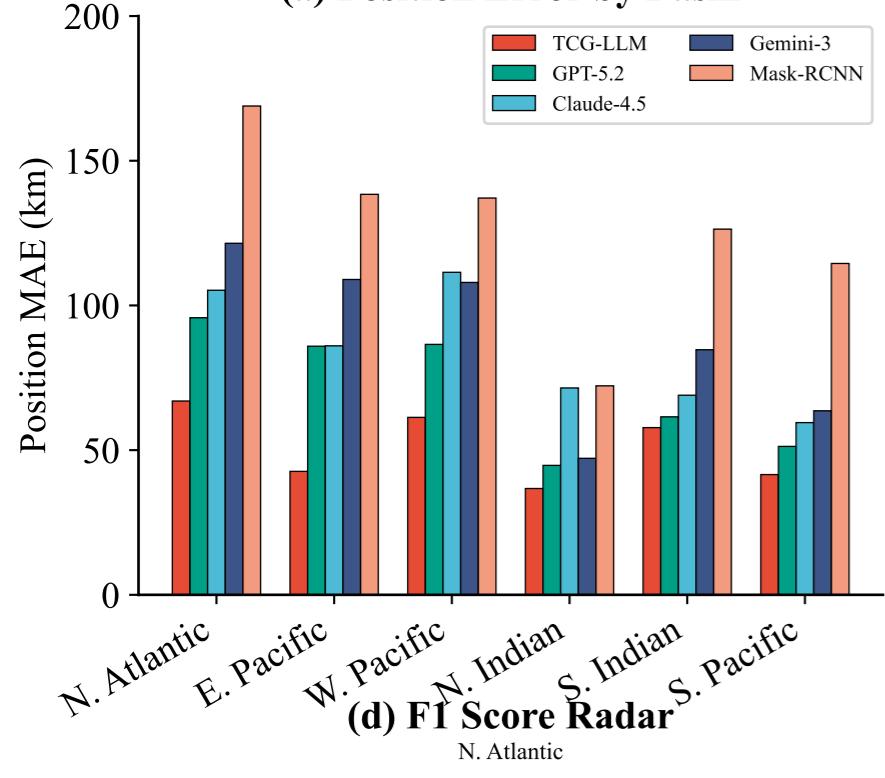
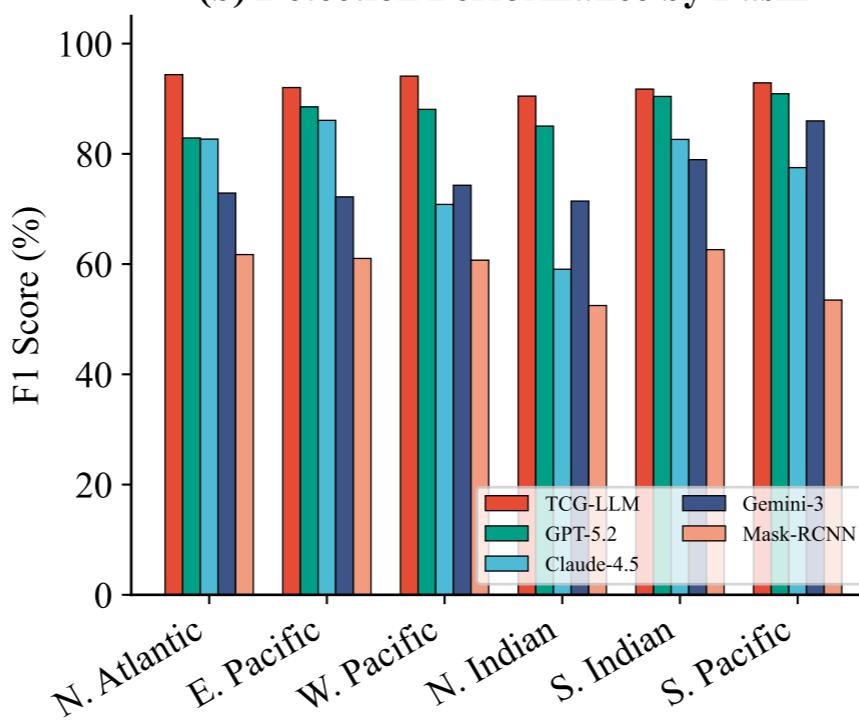


Multi-Model Basin-wise Performance Comparison for Tropical Cyclone Detection

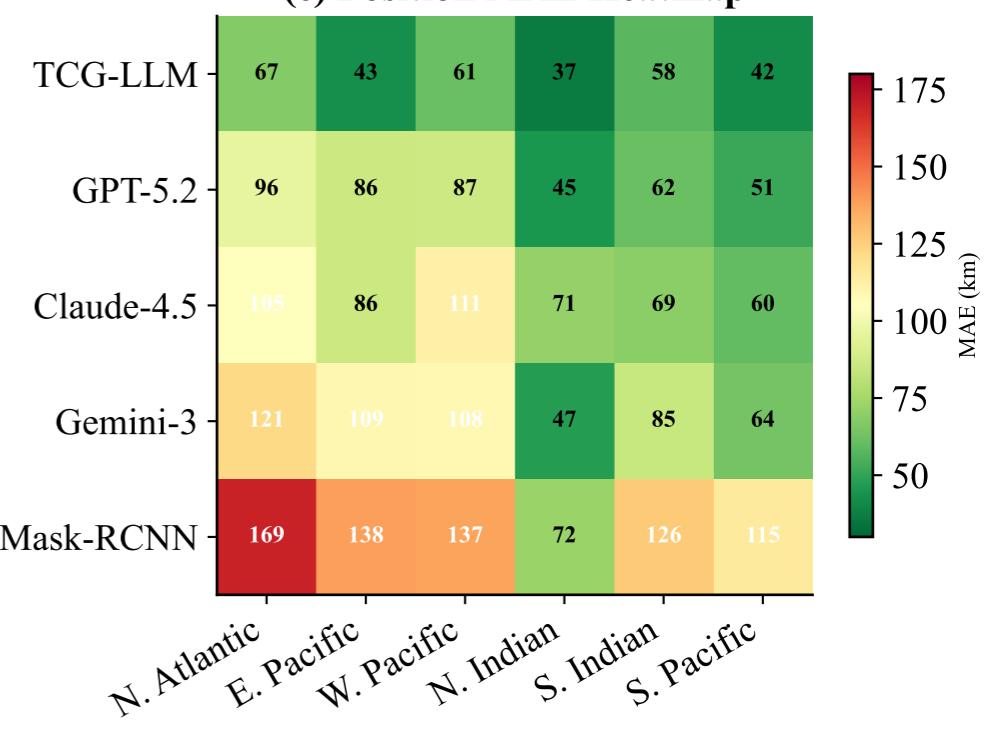
(a) Position Error by Basin



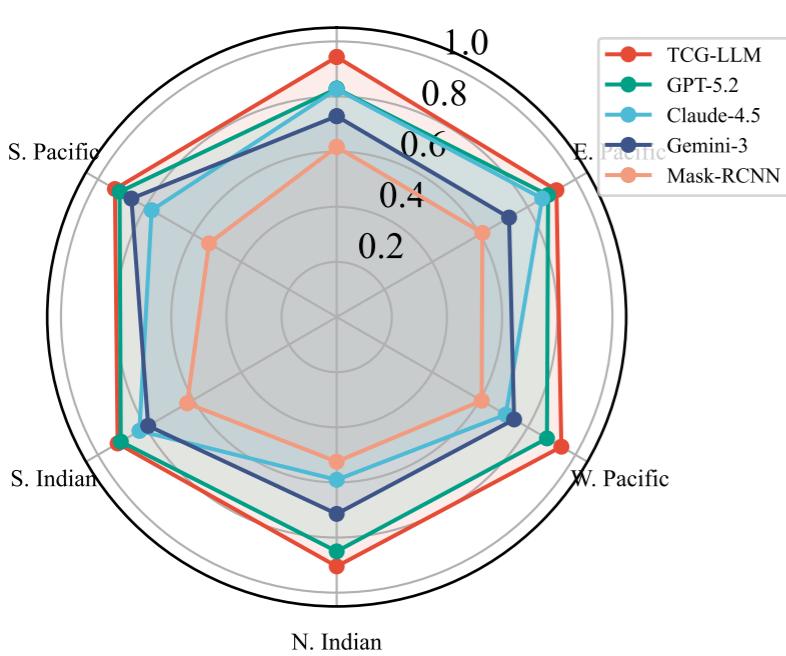
(b) Detection Performance by Basin



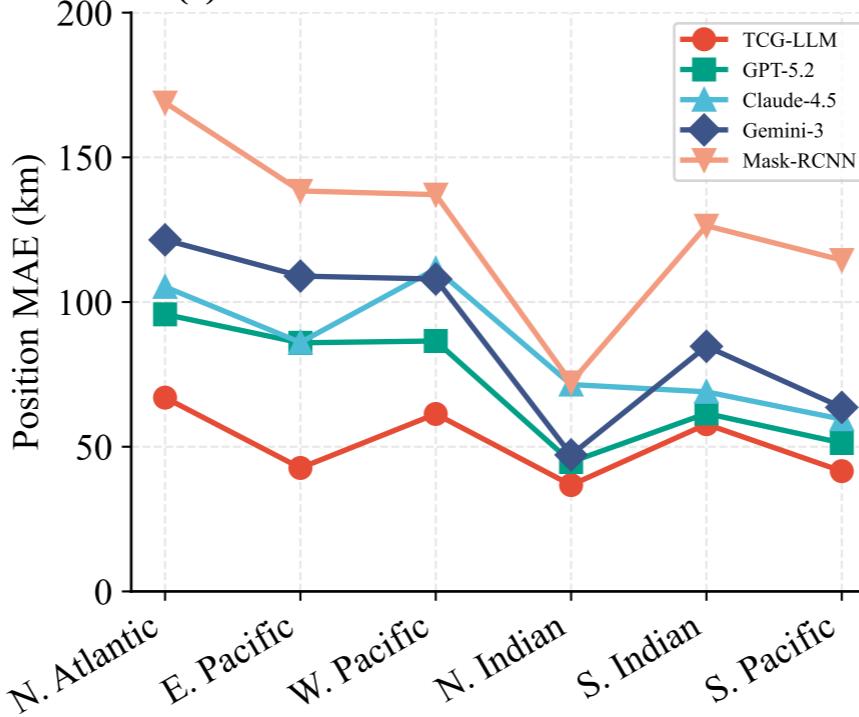
(c) Position MAE Heatmap



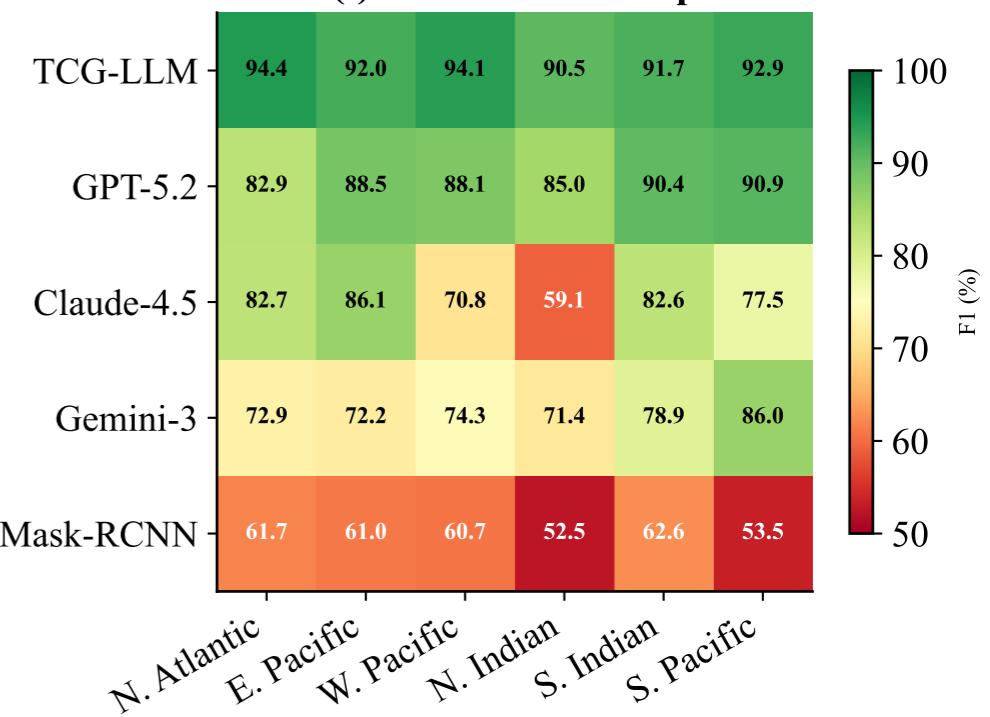
(d) F1 Score Radar



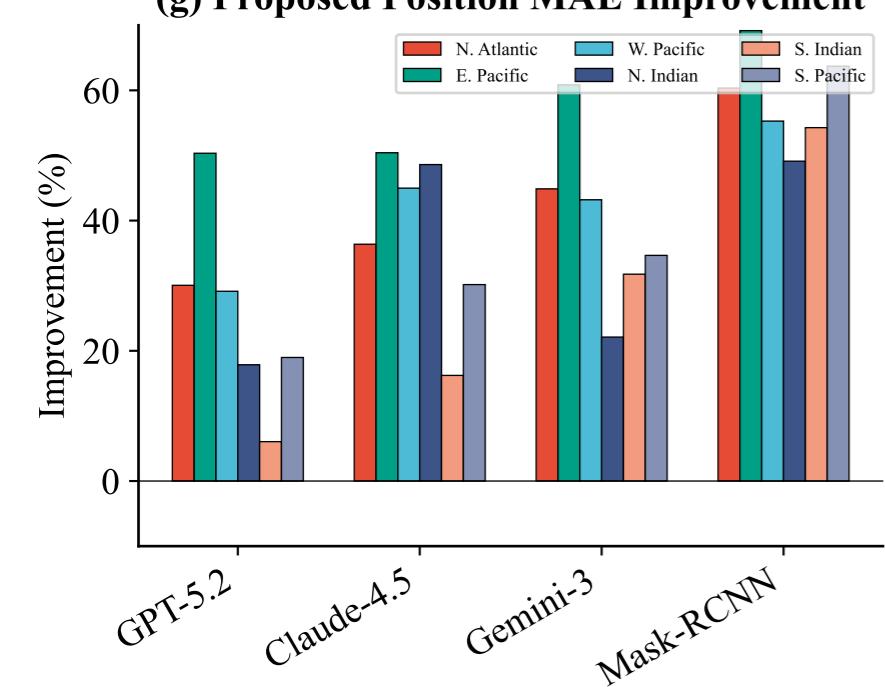
(e) Position Error Trend Across Basins



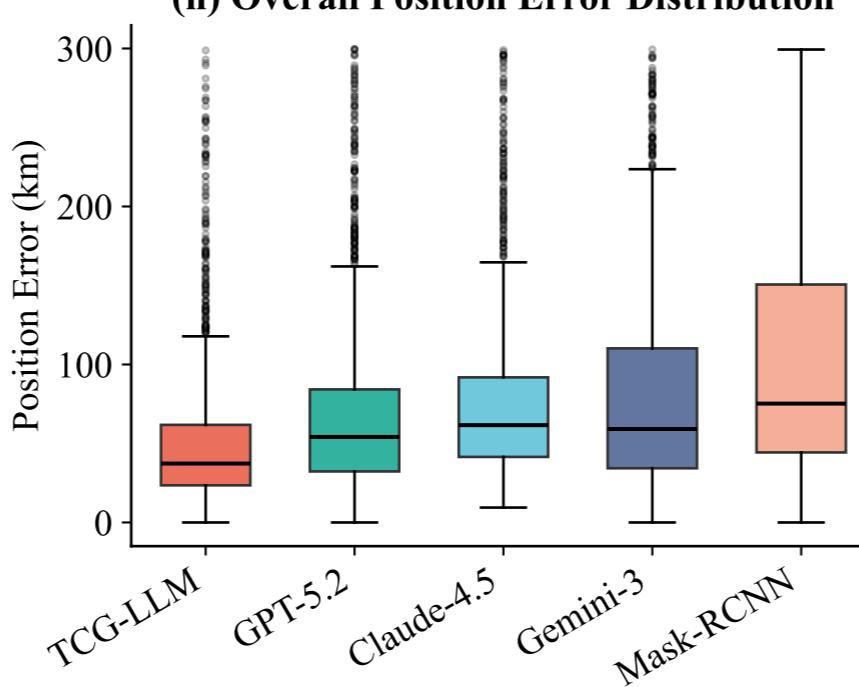
(f) F1 Score Heatmap



(g) Proposed Position MAE Improvement



(h) Overall Position Error Distribution



(i) Summary Statistics

Model	Avg MAE (km)	Avg F1 (%)	Best Basin	Worst Basin
TCG-LLM	51.2	92.6	N. Indian	N. Atlantic
GPT-5.2	71.0	87.6	N. Indian	N. Atlantic
Claude-4.5	83.8	76.5	S. Pacific	W. Pacific
Gemini-3	89.0	76.0	N. Indian	N. Atlantic
Mask-RCNN	126.3	58.7	N. Indian	N. Atlantic