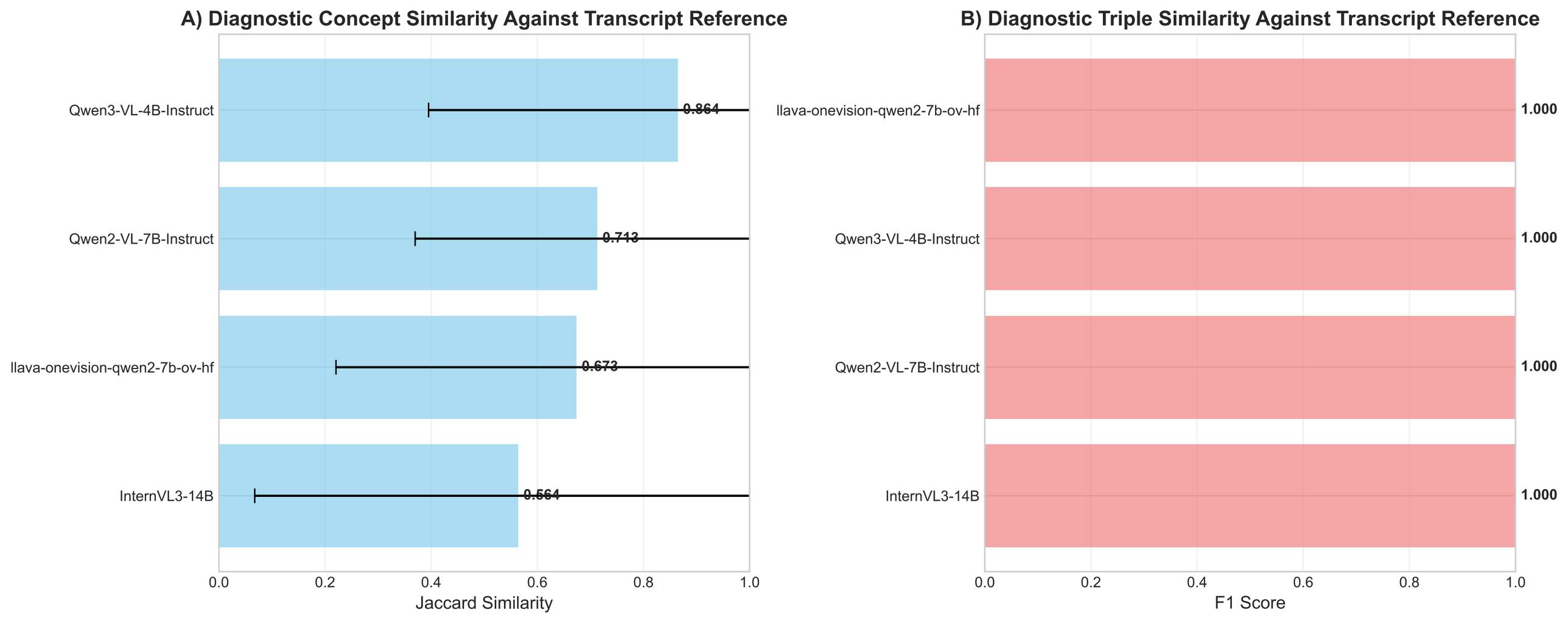
A) Diagnostic Concept Similarity Against Transcript Reference
Qwen3-VL-4B-Instruct
Qwen2-VL-7B-Instruct
llava-onevision-qwen2-7b-ov-hf
InternVL3-14B
0.864
0.713
0.673
0.564
0.0
0.2
0.4
0.6
0.8
1.0
Jaccard Similarity
B) Diagnostic Triple Similarity Against Transcript Reference
llava-onevision-qwen2-7b-ov-hf
Qwen3-VL-4B-Instruct
Qwen2-VL-7B-Instruct
InternVL3-14B
1.000
1.000
1.000
1.000
0.0
0.2
0.4
0.6
0.8
1.0
F1 Score