

Scenario A

Development

Dev Sample ID: 2812



0:00 6:45

Ground Truth ●



4:18 5:22

Prediction (Video-only) ●



5:01 5:21

Prediction (Heuristic Context) ●



4:19 5:22



- Ground truth (Dev) / Manual Inspection (Test)
- Prediction (Video-only)
- Prediction (VLM-Enhanced Context with Heuristic Context Guidance)

Test

Test ID: C64



0:00 6:16

Manual Inspection ●



3:42 3:51

Prediction (Video-only) ●



3:38 3:42

Prediction (Heuristic Context) ●



3:39 3:51



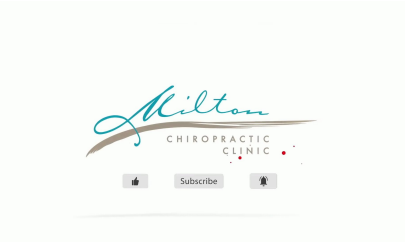
Scenario B & C

Development

Dev Sample ID: 2796



0:00



10:17

Ground Truth ●



8:04



9:25

Prediction (Zero-Shot with reasoning) ●



7:23



9:37

Prediction (Heuristic Context) ●



8:08



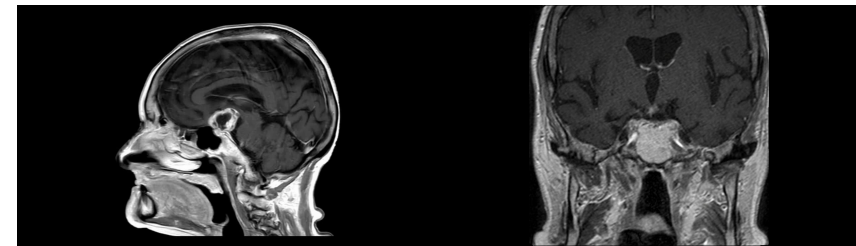
9:37



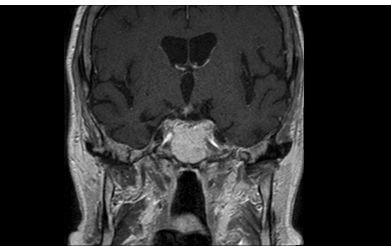
- Ground truth (Dev) / Manual Inspection (Test)
- Prediction (VLM-Enhanced Context with Zero-Shot with reasoning Guidance)
- Prediction (VLM-Enhanced Context with Heuristic Context Guidance)

Test

Test ID: C34

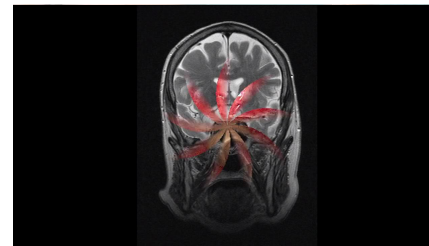


0:00

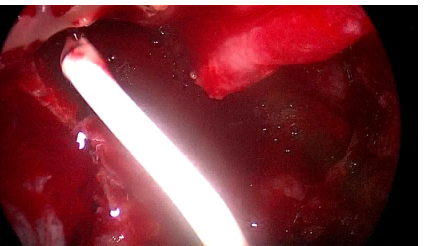


5:18

Manual Inspection ●

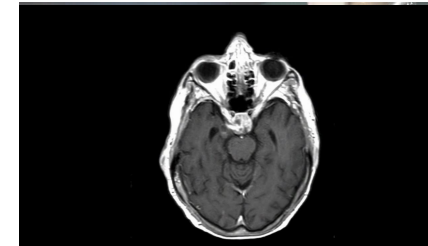


0:58

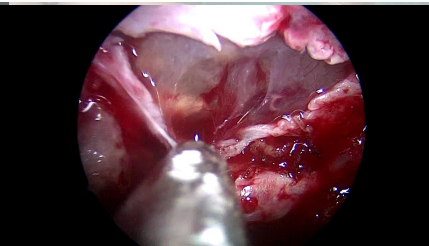


4:38

Prediction (Zero-Shot with reasoning) ●



0:39

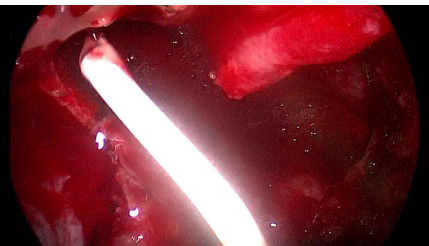


2:15

Prediction (Heuristic Context) ●



0:58



4:39

