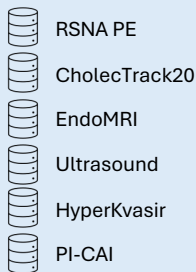


Public Cohort (n=216)

Source Datasets



Sampling Protocol

Has official test set

Randomly sample from official test set.

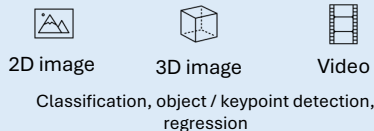
No official test set

Randomly sample cases with complete annotations and/or labels.

Inclusion Criteria

- Required modalities available
- Complete annotations / labels for the target tasks
- Readable image quality

Modalities & Tasks



Harmonization

Difficulty Assignment



Based on diseases, question types, modalities by heuristics.

Quality Check

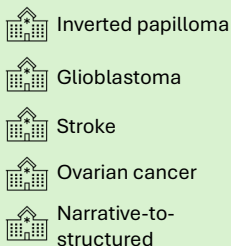
Verify the quality of annotations, consensus, formats, and de-identification.

Assembly

- Unify modalities, tasks, and clinical question types
- Standardize prompts
- Unify scoring metrics
- Ensure de-identification is HIPPA-compliant
- Standardize JSON formats for samples

Private Cohort (n=214)

Source Datasets



Information Extraction & Annotation



Consensus from 3 experts



Inclusion Criteria

- Relevant to target diseases/context
- Adequate data quality
- Availability of clinical information
- De-identifiable under IRB/HIPPA policy

Modalities & Tasks

