

Data Exploration of OmniMedVQA

Introduction (TODO)

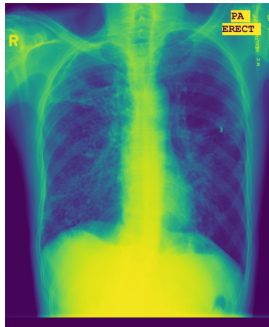
We are using the dataset OmniMedVQA which contains over 100,000+ images and question answer (qa) items from 73 different medical datasets [1]. Examining the dataset finds that the dataset contains [Todo: number of classes for question] classes of questions over all datasets.

Dataset Statistics (TODO)

- The types of problem types and number of classes
- The types of modalities and number of
- The number of images in each dataset
-

Visual Question Answering Examples

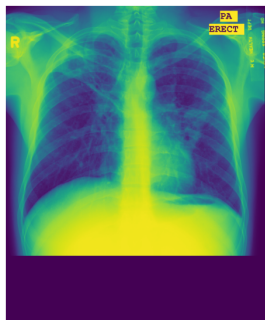
The two examples of visual question answering below are from the pulmonary Chest MC dataset.



Question:
What type of abnormality is present in this lung X-ray image?

Options:
A: Pleural effusion
B: Asthma exacerbation
C: manifestation of tuberculosis
D: Lung cancer

GT Answer: manifestation of tuberculosis
Modality: X-Ray



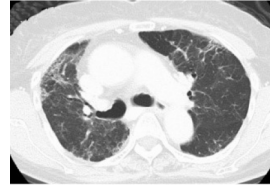
Question:
What is the imaging technique employed for capturing this picture?

Options:
A: Blood test
B: X-ray
C: PET scan
D: Ultrasound

GT Answer: X-ray
Modality: X-Ray

Within each dataset, there is a diversity in the types of questions being asked for a given image. The first image presents a modality question, which asks about the imaging technique used (such as X-ray, CT, or MRI), while the second image is a disease diagnosis question. All VQA items include multiple options relevant to the question posed, the ground truth, and the modality for the image.

This example is from a covid imaging dataset.



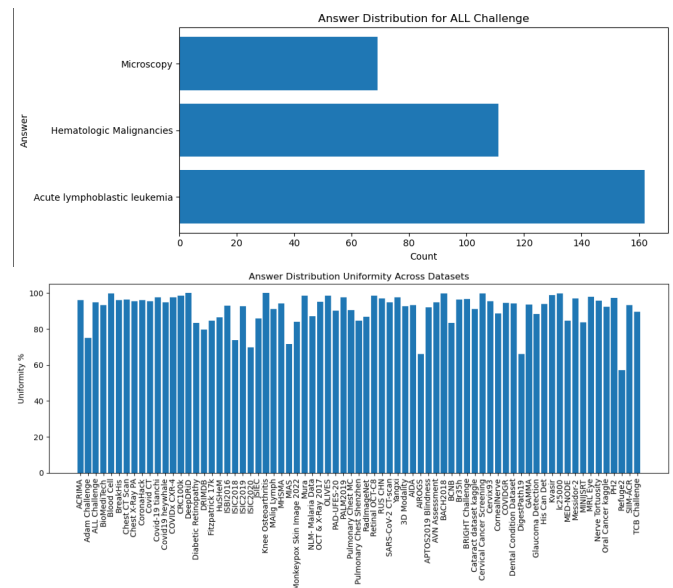
Question:
What type of imaging was employed to obtain this image?

Options:
A: Angiography
B: MRI
C: PET scan
D: CT.

GT Answer: CT.
Modality: CT(Computed Tomography)

Different modalities for imaging are used across all datasets but the same problem type can appear across different datasets. Questions are framed similarly across datasets creating consistency in the OmniMedVQA dataset. We also see that the quality of imaging can vary between datasets. Some of the images are high quality images using technology such as X-ray or a CT scan. But some of the images are camera pictures of human subjects with some background of the environment. In this example it is a picture taken of the output from a microscope. From the examples seen the text length of questions are short and direct with answer options being just a few words.

Challenges



Many of the questions show highly skewed answer distributions, with one option dominating while others are rarely chosen. This strong answer bias makes the data harder to interpret and can lead models to overfit to the majority choice rather than learning meaningful patterns.

