Breast Cancer Tumor Interpreter Report

Yuning Hu

Problem Definition

One of the primary applications of computer vision in biomedical contexts is identifying the location of tumors in X-ray images, with the goal of enabling affordable cancer screenings and assisting patients who may not have access to human pathologists. However, to truly simulate the role of a human pathologist, the system must also be capable of effectively communicating with patients. This includes not only locating tumors in images but also interpreting and explaining the results in an understandable way.

Breast cancer presents a particularly challenging case for interpretation. Standard screenings involve two mammographic views: the mediolateral oblique (MLO) and the craniocaudal (CC) views. Simply identifying the tumor's location on these images is insufficient, as patients generally lack the expertise to interpret mammograms, even when tumors are labeled. Human pathologists are trained to provide spatial descriptions of tumor locations using “clock positions,” which help patients better understand the findings. We aim for the automated agent to replicate this capability.

Given the MLO and CC mammograms, along with information about whether the image is of the left or right breast, the agent imagines the breast as a clock face and interprets the tumor’s position in terms of clock coordinates. In this project, we aim to design a data pipeline that enables extraction of the tumor’s clock position and supports the performance evaluation of different models or prompts.

Challenges and Solutions

Cross-Image Reasoning

Interpreting the clock position requires the agent to take both mammographic views as input and reason across them. The breast can be divided into four quadrants using conceptual x- and y-axes. The MLO (side) view helps determine whether the tumor is above, below, or aligned with the x-axis, while the CC (top-down) view helps determine whether the tumor is in the inner or outer part of the breast relative to the y-axis. The agent can only infer the quadrant in which the tumor is located by combining these two perspectives. This challenge can be mitigated by running experiments with models that allow multi-image input and have better performance on reasoning.

Breast Anatomy Distortion

Due to gravity, breast tissue naturally drapes downward. In the MLO (side) view, the x-axis does not correspond to a simple horizontal line extending perpendicularly from the nipple to the chest wall. Instead, the axis may be skewed. Accordingly, the clock positions would need to be evaluated with the non-standard quadrant. The agent should account for this anatomical distortion to accurately determine position. This challenge can be mitigated by including guidance in the prompt on how to identify the skewed x-axis.

X-ray Distortion

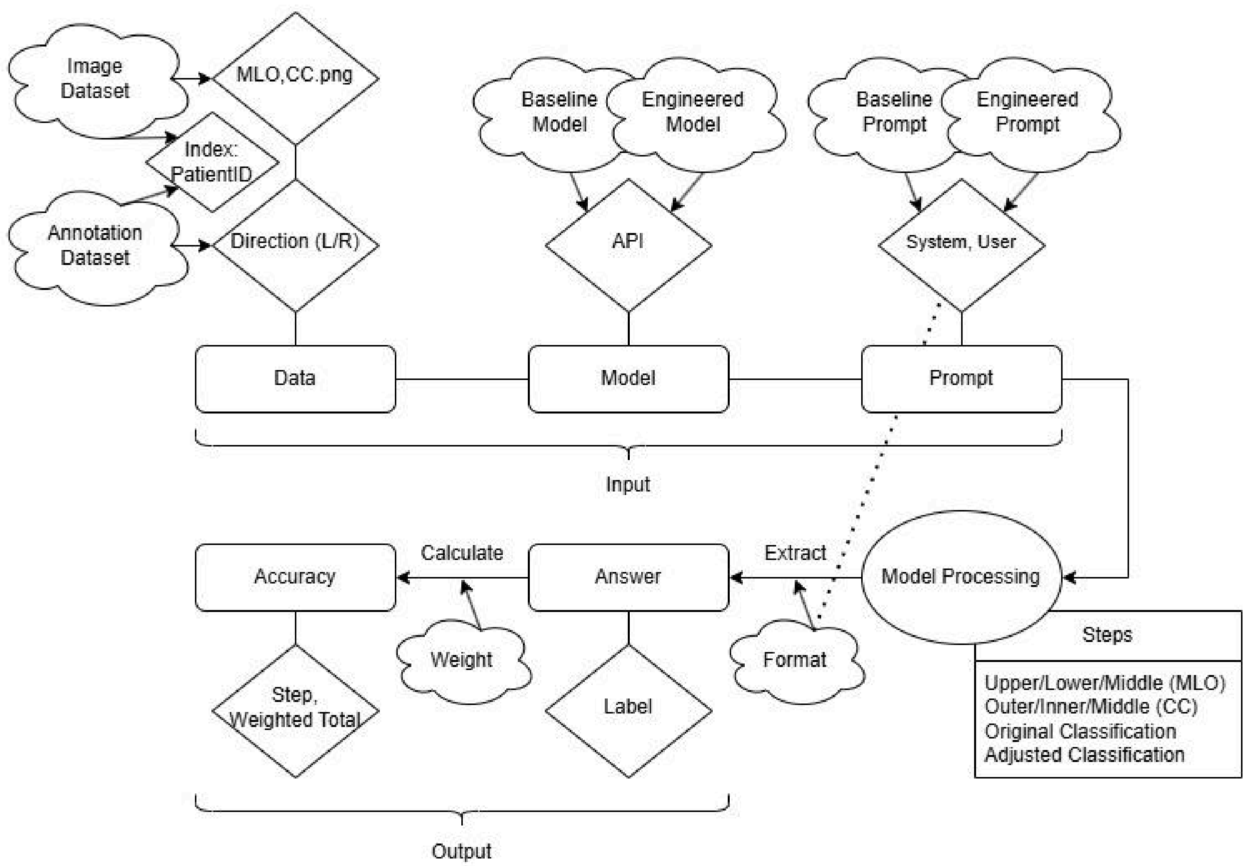
During mammography, the breast is partially compressed and distorted by the X-ray machine. The agent should account for this imaging distortion. This challenge can be mitigated by incorporating heuristics such as “Medial Up, Lateral Down” (MULD) or adjusted variants into the prompt.

Accuracy Assessment

To effectively assess and further improve model or prompt performance, accuracy evaluation must go beyond a simple comparison of predicted and actual clock positions. This challenge can be mitigated by computing accuracy at each step of decision making. Those steps are: Determining whether the tumor is upper or lower; Determining whether the tumor is inner or outer; Deriving the clock position, Adjusting for distortion.

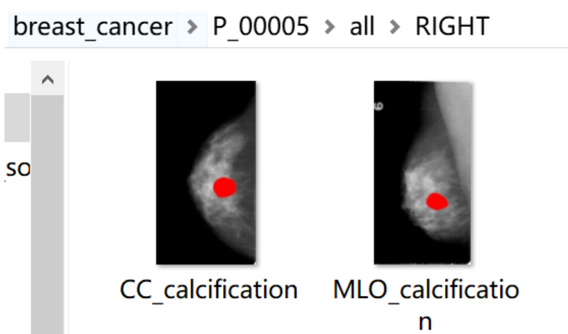
Breast Tumor Locator Pipeline

The pipeline design of a breast tumor locator. Blobs - can be customized. Arrows - from the source to the destination. Dashed lines - related.



Experimental Setup

The image dataset was curated and organized from the CBIS-DDSM (Curated Breast Imaging Subset of DDSM) dataset, as introduced in *A curated mammography data set for use in computer-aided detection and diagnosis research* (Nature). The annotation dataset was manually annotated after the *How to Read a Mammogram* (tutorial).



The baseline prompt was adapted from DeepSeek's summarization of the tutorial transcript. A comparative prompt was manually revised by a human annotator to include more detailed descriptions of clock-face division and anatomical adjustments.

The baseline model used was Qwen-VL-Plus-Latest, and a comparative model was Qwen-VL-Max-Latest.

Results

Based on the accuracy, Qwen-VL-Max-Latest consistently outperformed Qwen-VL-Plus-Latest. However, no significant difference in performance was observed between the two prompts.

Discussion

Fine-tuning: The model may benefit from fine-tuning with a small number (~10) of high-quality and representative annotated examples.

Input variation: Future experiments could test model performance on inputs of varying difficulty, such as mammograms with pre-drawn axes to aid spatial interpretation.

Annotation quality: The current performance may be underestimated, as the annotation dataset was manually labeled based on a lenient and somewhat subjective interpretation of the tutorial, rather than a strict, standardized heuristic.

Reflection

I found myself more toward designing data pipelines compared to fine tuning models.