

Overview

The rapid advancement of artificial intelligence (AI) has opened new frontiers in various fields, including medicine. Our project focuses on developing an AI assistant specialized in pathology to assist pathologists with diagnostic tasks and improve patient care. We propose an AI chatbot fine-tuned with pathology data using the large multimodal model LLaVA.

Data Extraction

We have curated a dataset of approximately 120k figure-caption pairs from over 450 pathology textbooks to train our model. To extract these figure-caption pairs, we compared two algorithms: PDFFigCapX and PDFFigures 2.0. Although both algorithms performed strongly, PDFFigCapX was superior in generating a larger number of accurate figure-caption pairs, making it our preferred choice for this task.

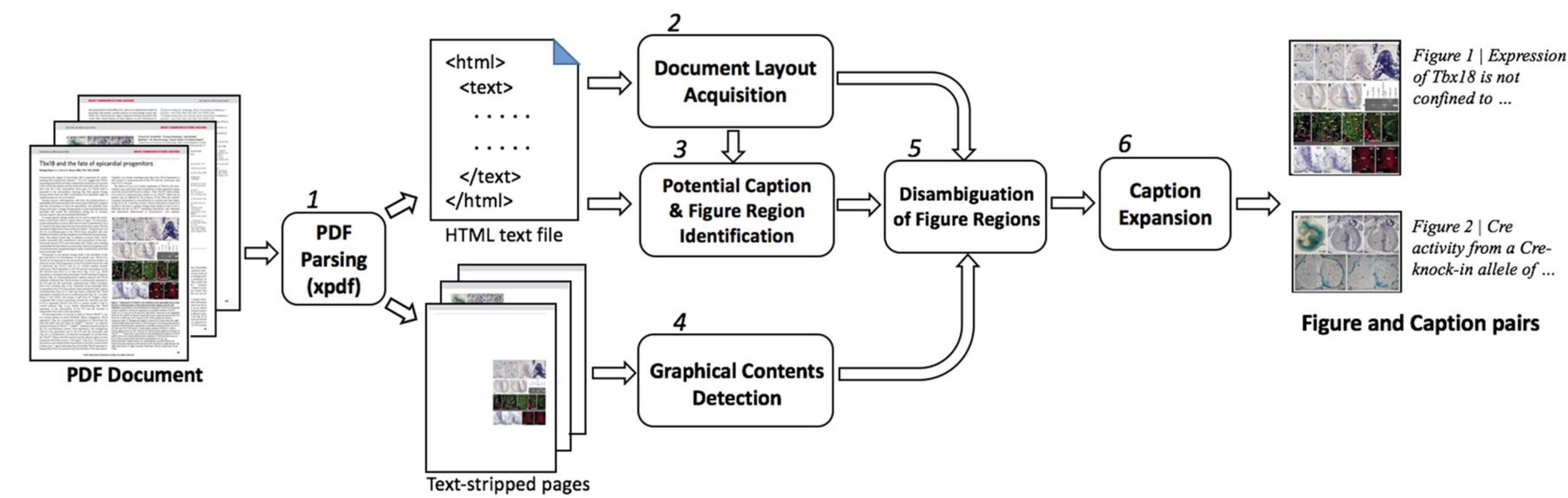
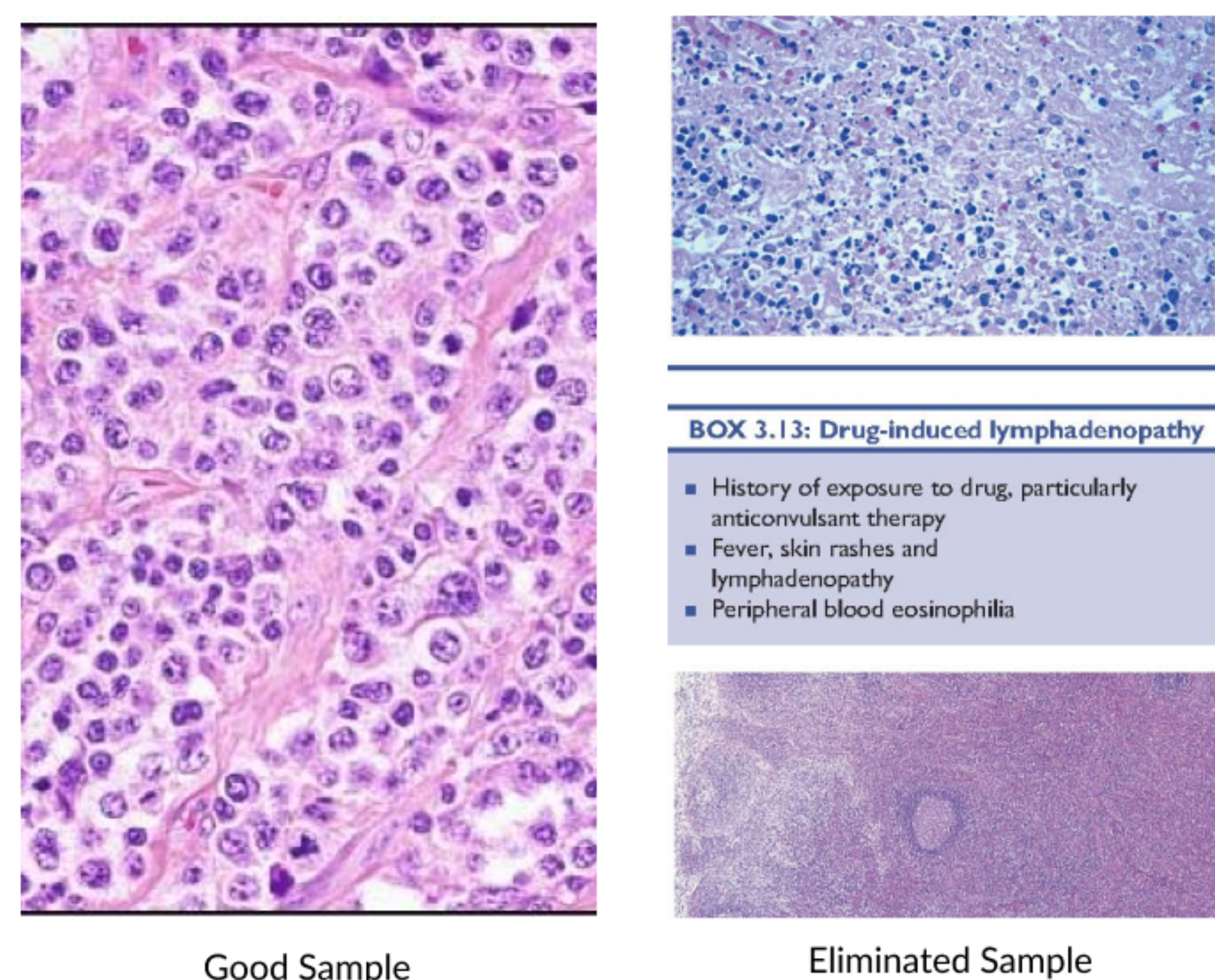


Figure 1. PDFFigCapX Extraction Process

Data Post-processing

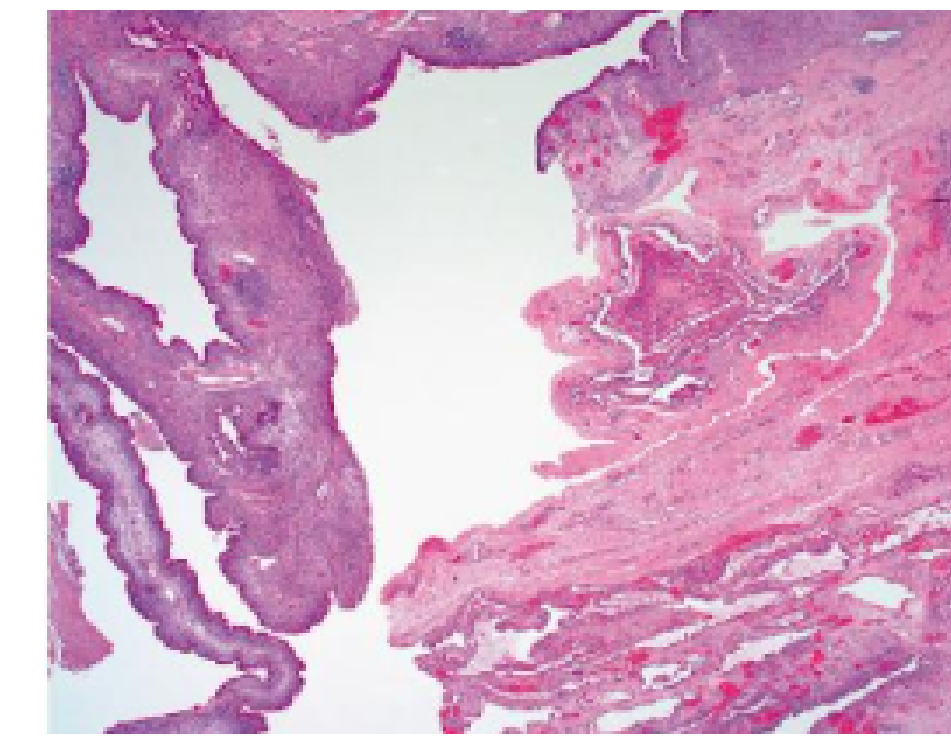
- Applied Optical Character Recognition (OCR) to identify and eliminate incorrect figure-caption pairs.
- Ensured the dataset's integrity and accuracy by confirming the presence of a corresponding valid pair.



Conversation Generation

- Based on our trials with various models: **GPT-4**, **Llama3** and **Gemini 1.0 Pro**, we have chosen to proceed with **Gemini 1.0 Pro**.
- Utilized Figure-Caption pairs to generate conversations, employing a carefully designed template.
- These conversations, along with the images, will serve as the training data.

Caption:
Figure 6-35: Low-power view of a type II pleuropulmonary blastoma. The spindle cell component is readily apparent.



(Image is shown only for reference)

Conversation:

Human: What type of cells are shown in the image?

Assistant: The image shows a low-power view of a type II pleuropulmonary blastoma. The spindle cell component is readily apparent.

Human: What is the significance of the spindle cell component?

Assistant: The spindle cell component is a characteristic feature of type II pleuropulmonary blastoma, a rare and aggressive type of lung cancer. Spindle cells are elongated, fibroblast-like cells that are often found in mesenchymal tissues, such as muscle, connective tissue, and blood vessels. In the context of pleuropulmonary blastoma, the presence of spindle cells suggests that the tumor may have a more aggressive behavior and a poorer prognosis.

Human: Can you describe the other components of the tumor?

Assistant: In addition to the spindle cell component, the tumor may also contain other components, such as epithelial cells, blastema, and cartilage.

Training

- Fine-tuned LLaVA using 4 x RTX 4090 24GB GPUs for ~3.5 hours, completing 1 epoch with ~120k figure-conversation pairs.
- Utilized DeepSpeed and LoRA to optimize memory usage and maintain performance efficiency.

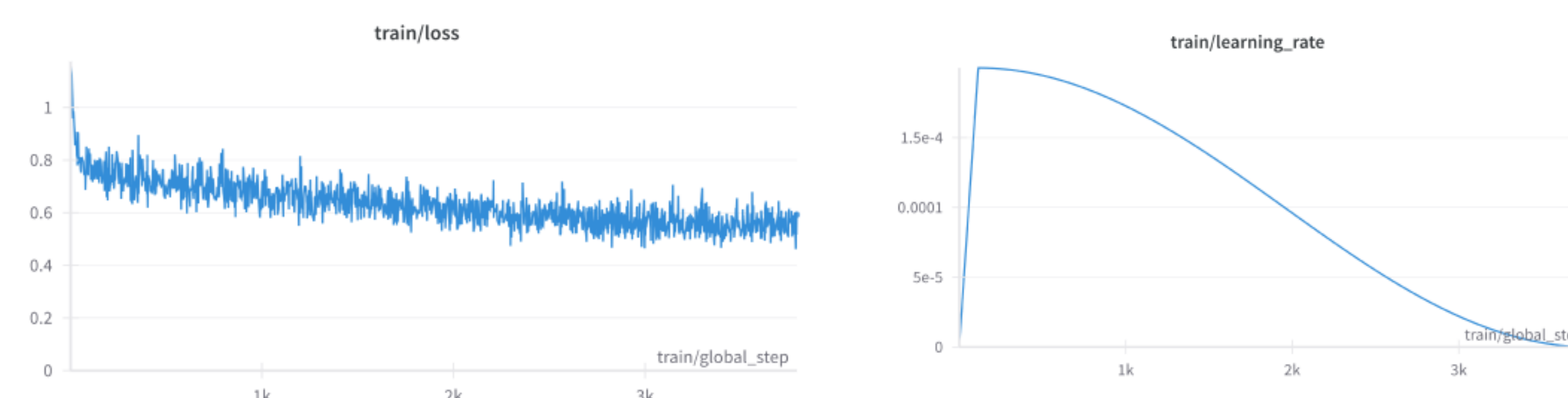
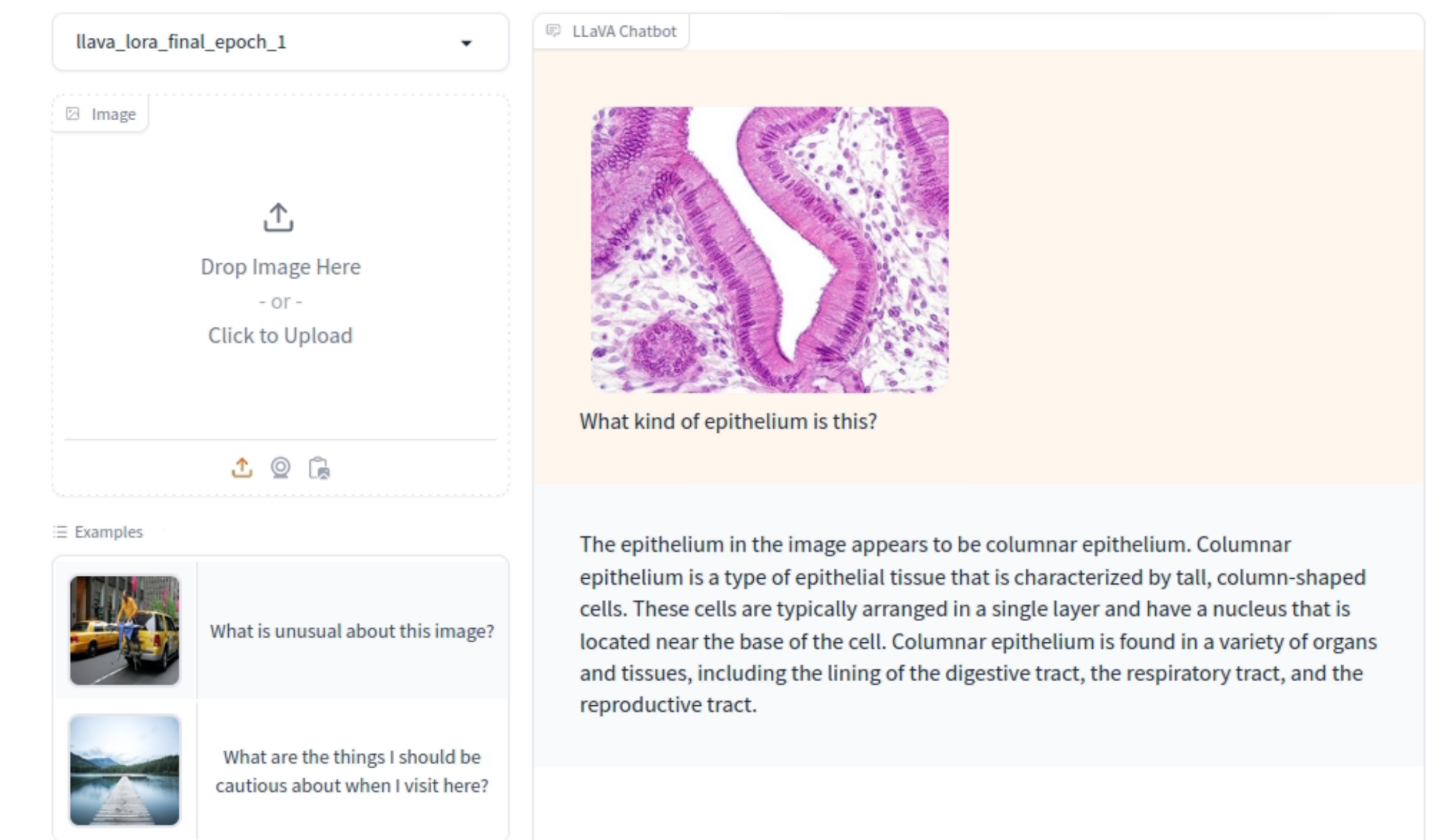


Figure 2. Loss and Learning Rate Plots

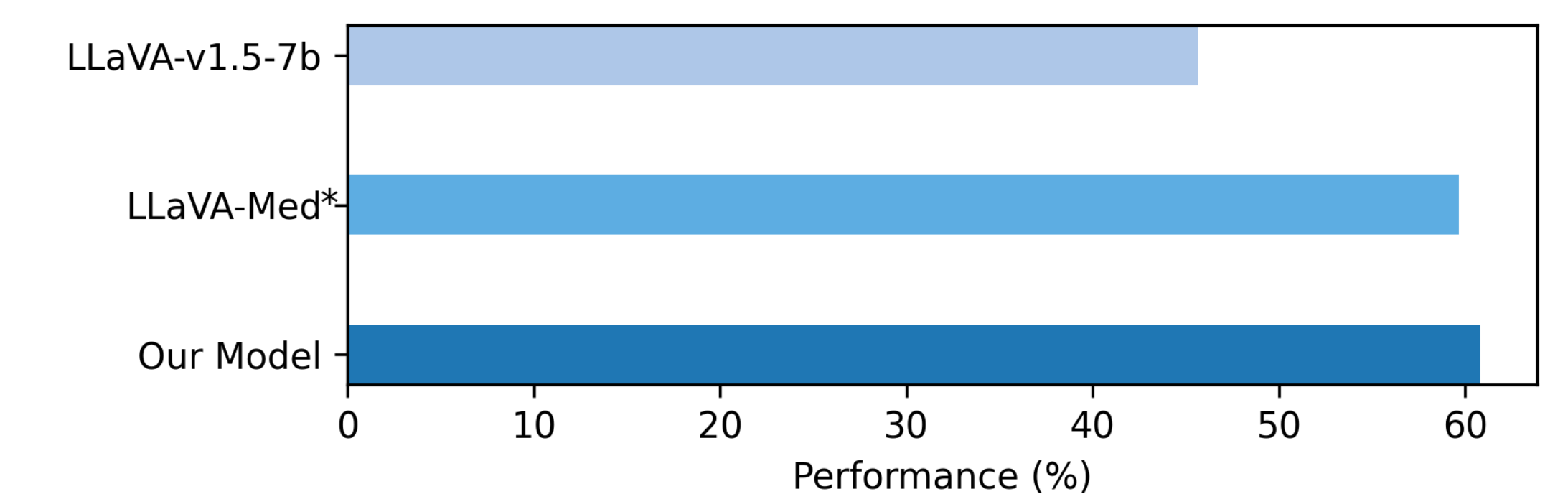
Inference

We utilized the LLaVA chatbot interface for ease of use, integrating it with our fine-tuned model for real-time pathology diagnosis.



Results

We evaluated our model's performance using PathVQA, a dataset comprising question-answer pairs based on pathology images. This dataset includes both open-ended questions and binary "yes/no" questions. For evaluation, we focused on the model's accuracy with the binary questions, testing over 3,250 such questions. We compared our results with the accuracy values reported in the LLaVA-Med paper for the LLaVA and LLaVA-Med models.



*The LLaVA-Med variant was trained on 600K samples for 1 epoch (stage 1) and 60K samples for 3 epochs (stage 2).

References

- Xuehai He, Yichen Zhang, Luntian Mou, Eric Xing, and Pengtao Xie. Pathvqa: 30000+ questions for medical visual question answering. *arXiv preprint arXiv:2003.10286*, 2020.
- Chunyuan Li, Cliff Wong, Sheng Zhang, Naoto Usuyama, Haotian Liu, Jianwei Yang, Tristan Naumann, Hoifung Poon, and Jianfeng Gao. Llava-med: Training a large language-and-vision assistant for biomedicine in one day, 2023.
- Pengyuan Li, Xiangying Jiang, and Hagit Shatkay. Figure and caption extraction from biomedical documents. *Bioinformatics*, 35(21):4381–4388, 04 2019.
- Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. Visual instruction tuning, 2023.