Dr-LLaVA: Visual Instruction Tuning with Symbolic Clinical Grounding

Shenghuan Sun* UCSF
Gregory M. Goldgof* MSK Cancer Center
Alexander Schubert UC Berkeley & UCSF
Zhiqing Sun CMU

Thomas Hartvigsen University of Virginia
Atul J. Butte UCSF
Ahmed Alaa UC Berkeley & UCSF

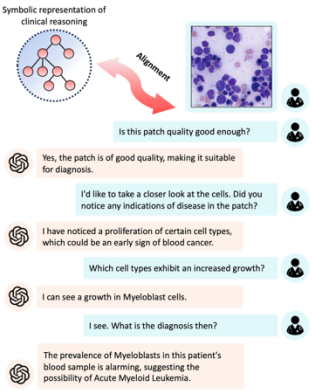

Figure 1: **Symbolic clinical grounding.** We utilize symbolic representations of diagnostic workflows to align multi-turn clinician-VLM conversations with valid clinical reasoning.

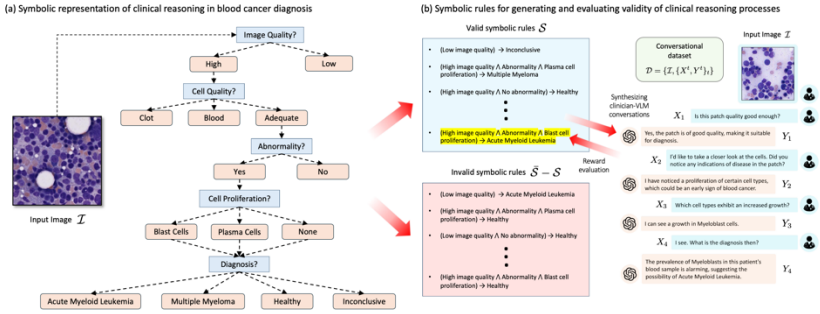

Figure 2: Depiction of our symbolic representation of clinical reasoning in blood cancer diagnosis.

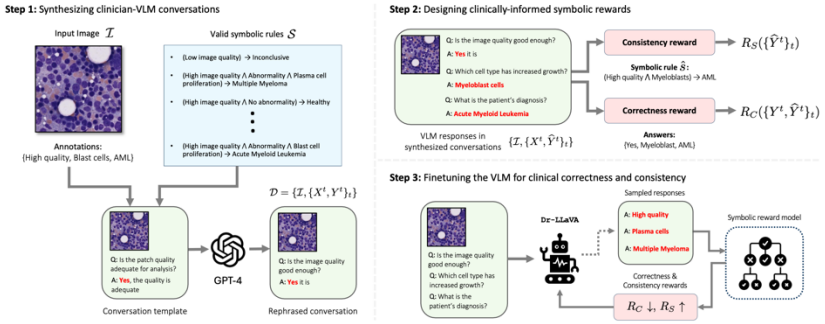

Figure 3: **Pictorial depiction of the `Dr-LLaVA` training pipeline** (a) For each medical image, we generate multi-turn conversations that are consistent with a symbolic model of clinical reasoning. This process uses GPT-4 to introduce diversity in the phrasing of the synthesized conversations. (b) We design a symbolic reward function that translates the responses provided by a VLM into symbolic rules and checking the correctness of the individual responses as well as the clinical validity of these rules. (c) Using the dataset in (a) and the reward model defined in (b), we use an RL approach in order to finetune a pretrained VLM.