SAGIVTECH IMVC 2026
15th Israel Machine Vision Conference (IMVC) 2026
Monday | April 27, 2026 | Pavilion 10, EXPO Tel Aviv

LVLM-Aware Multimodal Retrieval for RAG-Based Medical Diagnosis with General-Purpose Models

Nir Mazor, The Hebrew University of Jerusalem

Retrieving visual and textual information from medical literature and hospital records can enhance diagnostic accuracy for clinical image interpretation. However, multimodal retrieval-augmented diagnosis is highly challenging. We explore a lightweight mechanism for enhancing diagnostic performance of retrieval-augmented LVLMs. We train a lightweight LVLM-aware multimodal retriever, such that the retriever learns to return images and texts that guide the LVLM toward correct predictions. In our low-resource setting, we perform only lightweight fine-tuning with small amounts of data, and use only general-purpose backbone models, achieving competitive results in clinical classification and VQA tasks compared to medically pre-trained models with extensive training. In a novel analysis, we highlight a previously unexplored class of errors that we term inconsistent retrieval predictions: cases where different top-retrieved images yield different predictions for the same target. We find that these cases are challenging for all models, even for non-retrieval models, and that our retrieval optimization mechanism significantly improves these cases over standard RAG. However, our analysis also sheds light on gaps in the ability of LVLMs to utilize retrieved information for clinical predictions. Code and models available at: https://github.com/Nirmaz/CLARE.