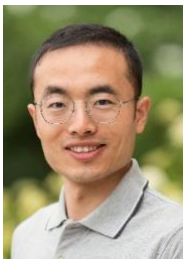
HBKU Thematic Research Grant 3rd Cycle– Project Highlight

Project Title: Interactive Multimodal AI for High-precision Medical Diagnosis



Zhihe Lu

Executive Summary (limit to 200 words)

The proposed project aims to advance medical artificial intelligence (AI) by developing an interactive multimodal AI system capable of performing high-precision diagnostic tasks guided by natural language instructions. This research addresses a critical challenge in AI-assisted medicine: the need for systems that not only process multimodal data (e.g., vision and text) but also interact with clinicians to perform complex visual reasoning tasks, such as identifying, segmenting, and counting medical structures from clinical images. Specifically, our research will focus on developing a model architecture that combines: (i) A visual encoder that extracts fine-grained features from medical imaging; (ii) A text encoder that processes clinician-provided language instructions; (iii) A multimodal fusion module that aligns and interprets both modalities to generate task-specific outputs (e.g., cell counts, object locations, or binary diagnoses); (iv) A feedback mechanism for real-time interaction and refinement of generated report. The immediate application scenario involves cell counting in microscopy images, a task critical to cancer diagnosis and treatment planning. We design datasets and evaluation protocols that simulate real-world human-AI interaction, enabling us to assess performance, interpretability, and usability.

Expected Outcome (limit to 100 words)

The expected outcomes of this project include: (i) A prototype of an interactive multimodal AI system for medical image understanding; (ii) Rigorous benchmarks and validation on clinical cell analysis datasets; (iii) Peer-reviewed publications and open-source contributions aligned with HBKU's thematic focus on AI and Precision Medicine.

Collaborating HBKU entities:

College of Science and Engineering (CSE) and Qatar Computing Research Institute (QCRI)

Photos – please insert photos, schematics, graphs…etc. relevant to the project

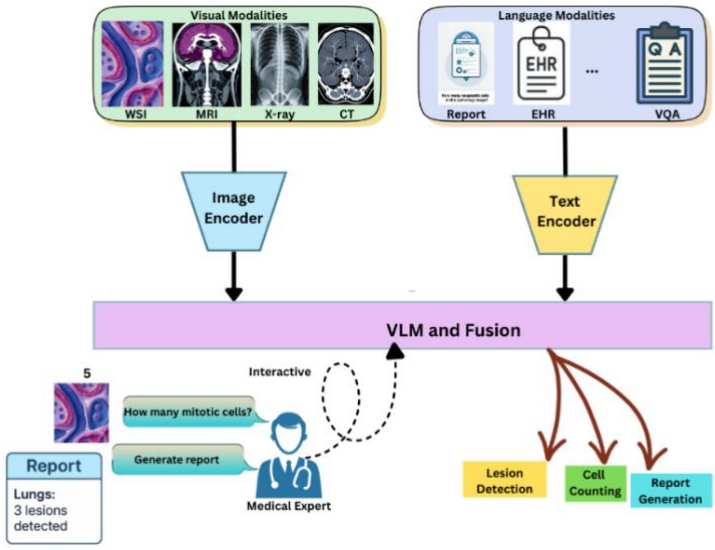


Figure 1 High-level conceptual diagram of the proposed Interactive Multimodal AI system

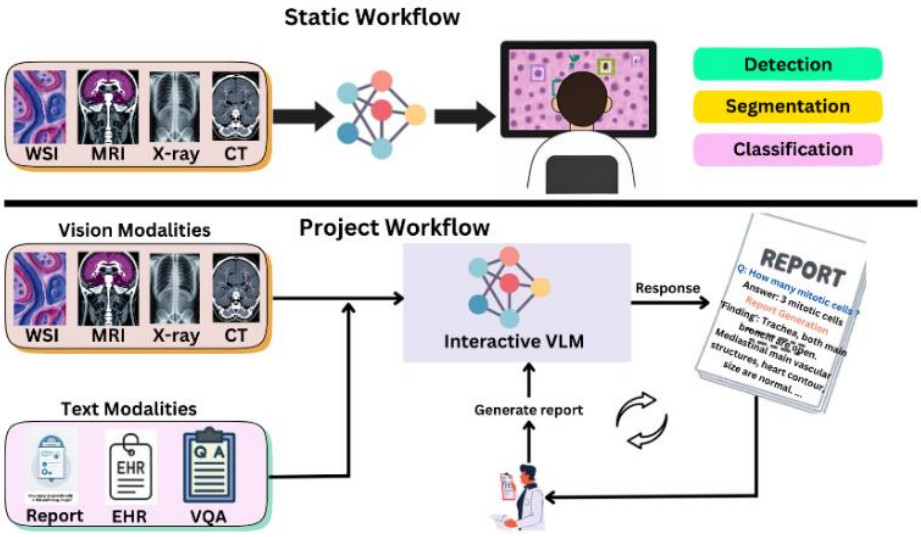


Figure 2 Comparison of current static AI workflows (top) with the proposed interactive multimodal workflow (bottom). Traditional systems remain task-specific, while the proposed design integrates visual and textual modalities, enabling interactive reasoning and report generation.