Agent-Based Multimodal Medical Reasoning

Research Paper Blueprint: Full Methodology, Experiments, and Figures

1. Research Question

Do multi-agent clinical reasoning systems (Radiologist + Clinician + Verifier) improve clinical reasoning reliability compared to monolithic multimodal models?

2. Motivation

Clinical decision-making is inherently collaborative. Radiologists interpret images, clinicians integrate patient history, and senior physicians verify conclusions. Modern multimodal models collapse this workflow into a single forward pass, potentially increasing hallucinations and silent failures.

3. Proposed Agents

Radiologist Agent: Interprets medical images and extracts visual findings.
Clinician Agent: Performs diagnostic reasoning using text and visual findings.
Verifier Agent: Checks consistency, evidence grounding, and revises conclusions.

4. Baseline Model

A monolithic multimodal model using image and text encoders followed by a fusion module and decoder. No explicit verification or correction stage.

5. Agent-Based Architecture

Image → Radiologist Agent → Findings → Clinician Agent → Diagnosis + Reasoning → Verifier Agent → Final Decision. This decouples perception, reasoning, and verification.

6. Datasets

VQA-RAD for medical visual question answering and MIMIC-CXR for radiology diagnosis and report reasoning.

7. Experiments

Compare monolithic vs agent-based systems across accuracy, error correction rate, contradiction detection, and confidence calibration. Perform ablation by removing agents.

8. Evaluation Metrics

Accuracy and F1 for correctness. Error Correction Rate measures how often verifier fixes errors. Contradiction Rate measures reasoning conflicts. Expected Calibration Error measures confidence reliability.

9. Expected Results

Agent-based models slightly improve accuracy but significantly reduce contradictions and overconfidence, demonstrating superior reliability.

10. Figures

Figure 1: Agent-based architecture diagram.
Figure 2: Contradiction rate comparison.
Figure 3: Calibration curves.
Figure 4: Qualitative case study.

11. Conclusion

Agent-based multimodal reasoning aligns better with real clinical workflows and offers a safer alternative to monolithic medical AI systems.