MEDIcal Question Answering For Multimodal And Generative telemedICine - MEDIQA MAGIC For Dermatology

Bhavana Anand, Charumathi P and Dr. S Kavitha,
SSN College Of Engineering, Chennai, India.

Task Description: Develop an AI-powered tool to generate automated, natural language dermatological responses based on textual health queries and accompanying images.

Objective: To develop an AI-driven system that generates accurate, context-aware dermatological responses from user queries and images.

Submission: Springer Journal of Intelligent Information Systems

Highlights

The proposed research aims to achieve the following:

Multimodal Input Integration: Enable **text and image-based** queries for context-aware dermatology QA.

Enhanced Response Generation: Use **LLMs** with **Chain of Thought(CoT)** reasoning for accurate and interpretable natural language responses.

Factually Accurate Knowledge Base: Integrate a **multimodal** Knowledge Graph(KG) for precise & context-aware responses.

Identified Research Gaps

1. Dataset Limitations: A lack of dermatologist-annotated medical images **limits model performance**. The MEDIQA-M3G dataset, with only 350 annotated responses, is **insufficient** for robust training.

2. Misdiagnosis of Diseases: Existing models often misdiagnose(rely on fine-tuned LLMs and zero-shot inference), leading to **hallucinations** and **unreliable** responses.

3. Lack of Interpretability: Current systems fail to provide transparent reasoning, making responses **unaccountable** due to **weak reasoning** capabilities.

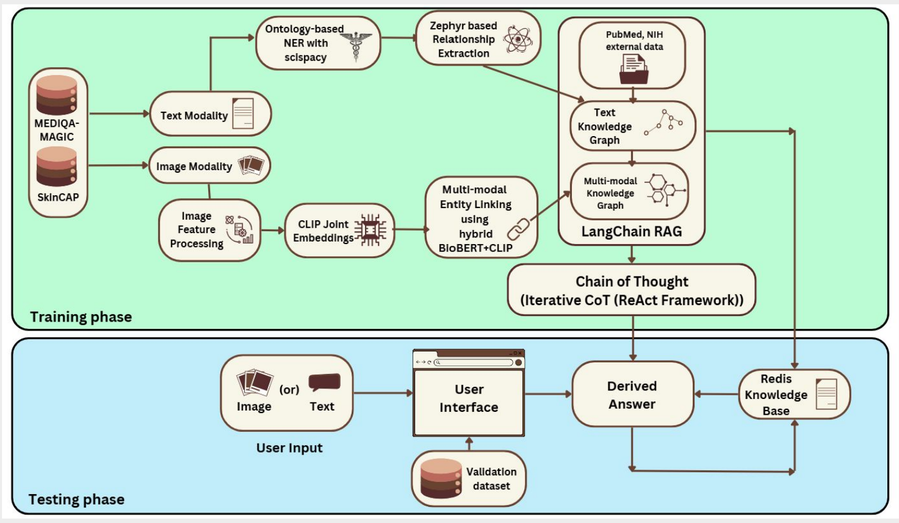
Data set Description

Dataset	No. of Annotated Data
MEDIQA-MAGIC	265
SkinCAP	4000

Dataset Summary

Image	Query	Response
10cx1if 	this spread of my hand to my lip and face it come and go it painful itch swollen and it look like tiny blisterspimples please help	I need to know if it is warm and hard to touch, is it spreading? Have you applied something after which it happened? It can be dermatitis or infection, though it looks more like dermatitis so starting the treatment for same. Wash the face with mild soap, apply moisturizer. Use sun protection. Apply the cream which i have added twice daily for 2 weeks. Take an antihistamine and painkiler on an as needed basis. Take tab for 5 days in the morning after breakfast.

Methodology



- Data processing is performed using ontology-based **Named Entity Recognition** to extract medical entities.
- **Entity relationships** are extracted using Zephyr-7b to link text entities with image features and construct a text-based **Knowledge Graph** representing medical concepts as triplets.
- **Multimodal embeddings** are generated using a fine-tuned CLIP model to align text and image modalities to form a multimodal knowledge graph.
- Medical knowledge retrieval is enabled through a **multimodal and graph-based Retrieval-Augmented Generation** module.
- **Session-based iterative Chain-of-Thought reasoning** is applied for coherent and logically structured responses.
- An interactive system presents responses effectively for telemedicine support.

Performance

Metric	Without KG	Text KG	Multimodal KG
BERT Score	0.739	0.812	0.859
Contextual Rel	0.55	0.72	0.74
Groundedness	0.65	0.77	0.81

Evaluation metrics for configurations :

Without knowledge graph(KG), with knowledge graph, and with multimodal knowledge graph

Model Name	Methodology	BERT Score
Proposed System	Multimodal Knowledge Graph and Graph RAG	**0.867**
VisionQAaries [6]	Fine-tuning Small Vision Language Models (Moondream, TinyLLaVA)	0.848
IRLab@IIT_BHU [1]	Generation-based LLM and Classification with BiLSTM + MLP	0.839
YuanAI [8]	Parameter-efficient fine-tuning of Llama3 and Gemini	0.835

Comparison of BERT Scores across existing models

Inferences

- Improved Accuracy with Multimodal KG and Reasoning – The integration of **KG, CoT Reasoning, and fine-tuned CLIP embeddings** significantly enhances the accuracy and contextual understanding of telemedicine responses, outperforming baseline models with a **6% improvement** in BERT Score.
- Scalability and Future Research Potential – The **dynamic KGs in a Graph RAG pipeline** offer real-time adaptability to evolving medical data, making the system scalable to **other medical domains** beyond dermatology for enhanced fact-based, interpretable medical responses.