

Multi-Modal Medical Assistant for Skin Diseases

(Ansh Lulla, Advik Sharma)

Overview

The Multi-Modal Medical Assistant integrates **Visual Retrieval-Augmented Generation (VIS-RAG)** and **Text Retrieval-Augmented Generation (Text-RAG)** to analyze multi-modal medical data for classification tasks. It leverages **LangSmith** for evaluation, monitoring, and debugging, ensuring efficient and accurate decision-making.

VIS-RAG: Visual Data Pipeline

1. **Input:** Image of the Skin Disease.
 2. **Embedding Model:** Encodes image features into vector embeddings for similarity analysis.
 3. **Image Index:** Stores embeddings for retrieval.
 4. **Metadata Annotations:** Adds labels and contextual information.
 5. **Generator:** Produces outputs, such as diagnostic insights or annotated images.
 6. **Output:** Image-based diagnostic results or features for decision-making.
-

Text-RAG: Text Data Pipeline

1. **Input:** Textual data (e.g., symptoms, reports).
2. **Text Database:** Retrieves relevant medical knowledge using embeddings.
3. **Conversational Augmentation:** Dynamically queries users for additional context to refine inputs.

4. **Generator:** Processes inputs and retrieved data for detailed analysis.
 5. **Output:** Text-based diagnostic results or augmented insights.
-

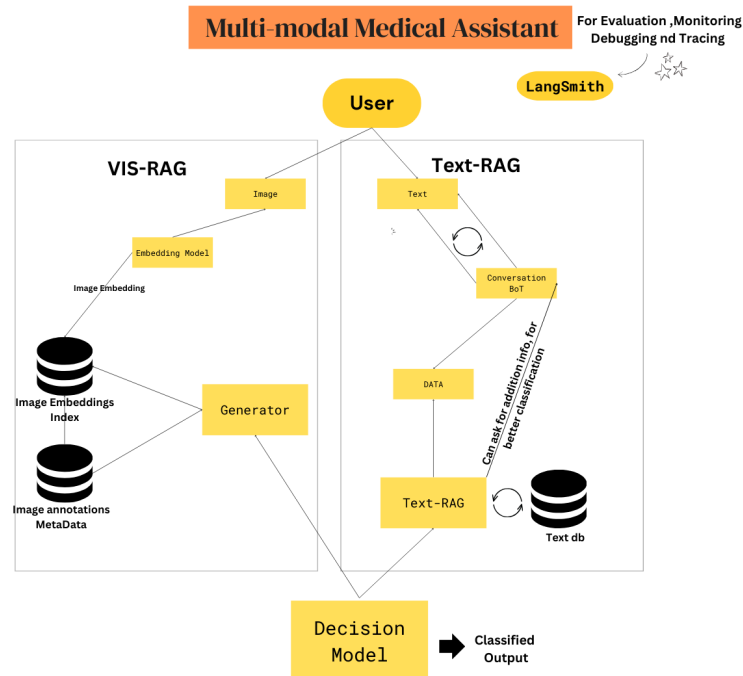
Integration Workflow

1. VIS-RAG and Text-RAG independently process visual and textual inputs.
 2. Outputs from both pipelines are combined via a **Decision Model**, generating a comprehensive, classified result.
 3. LangSmith tools trace data flow, ensuring transparency and debuggability.
-

Benefits

1. **Multi-Modal Retrieval:** Combines visual and textual embeddings for enriched analysis.
2. **Dynamic Augmentation:** Enhances input quality through iterative clarifications.
3. **LangSmith Monitoring:** Tracks data flow for debugging and evaluation.
4. **Efficient Embedding Storage:** Utilizes vector indexing for fast, scalable retrieval.
5. **Unified Decision Output:** Delivers actionable, multi-modal insights.

Diagrammatic Representation



Applications

1. Healthcare decisions often require correlating visual data (e.g., medical imaging) with textual data (e.g., patient records, clinical notes). The integration of VIS-RAG and Text-RAG enables enhanced decision-making through the Multi-Modal Data Integration.
2. For radiology, dermatology, and pathology, VIS-RAG processes image data with high precision, supporting accurate diagnoses.
3. Text-RAG efficiently processes the symptoms, and patient histories, providing valuable insights for decision support and efficient diagnosis.

Conclusion

The Multi-Modal Medical Assistant enhances healthcare facilities by integrating visual and textual data analysis, enabling accurate diagnostics and personalized care. Its dynamic retrieval and multi-modal fusion ensure efficiency and adaptability across diverse medical applications. This system thus represents a significant step towards leveraging Generative AI for improved patient outcomes and advanced medical research.