MedAI: Multimodal Clinical Decision Support System

Abstract

This project develops a comprehensive Multimodal Medical Assistant integrating medical imaging analysis with clinical text processing. The system processes medical images and prescription texts using dual AI models - Google Gemini 2.0 Flash for primary analysis and Llama 3.1 via Groq API for clinical verification. By combining computer vision with NLP, the assistant provides integrated clinical assessments, anomaly detection, and treatment recommendations. The implementation handles five medical cases, generating structured JSON outputs correlating imaging findings with prescription data. Key outcomes include successful multimodal data fusion, accurate medical entity extraction, and clinically relevant analysis validated through dual LLM verification.

1. Introduction

Context and Background

Healthcare faces challenges managing complex medical data across multiple modalities. Traditional systems process data in isolation, missing critical correlations between imaging findings and clinical context. Multimodal AI systems offer transformative potential by integrating diverse medical data sources.

Motivation

This project addresses the need for integrated clinical decision support systems that process both visual and textual medical data simultaneously, bridging the gap between radiological findings and clinical documentation.

2. Problem Statement

Healthcare providers work with disconnected systems leading to:

- Fragmented patient information across modalities
- Missed correlations between imaging and clinical symptoms
- Increased diagnostic time due to manual data correlation
- Limited integration between prescription data and diagnostic imaging

3. Objectives

- Develop multimodal system for medical images and prescription texts
- Implement dual LLM architecture for clinical analysis
- Create medical image analysis pipeline
- Build clinical text processing system

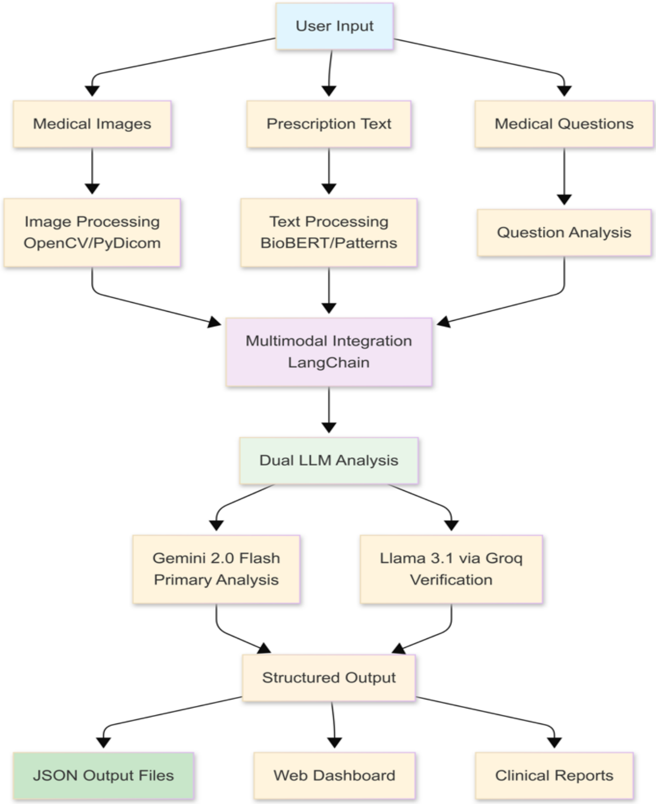
- Generate structured JSON outputs
- Develop user-friendly interface

4. Methodology

- **Tools and Technologies**
  - LLMs: Gemini 2.0 Flash, Llama 3.1 via Groq API
  - Medical Vision: CLIP with medical-optimized prompts
  - Clinical Text: BioBERT for entity recognition
  - Medical Imaging: OpenCV, PyDicom
  - Orchestration: LangChain
  - Interface: Streamlit
- **Workflow**
  - Data Ingestion: Medical images + prescription texts
  - Feature Extraction: Image analysis + text processing
  - Dual LLM Analysis: Gemini + Llama verification
  - Multimodal Integration: Image-text correlation
  - Output Generation: Structured JSON reports

5. System Design / Implementation

- **Architecture diagram**


User Input
Medical Images
Prescription Text
Medical Questions
Image Processing OpenCV/PyDicom
Text Processing BioBERT/Patterns
Question Analysis
Multimodal Integration LangChain
Dual LLM Analysis
Gemini 2.0 Flash Primary Analysis
Llama 3.1 via Groq Verification
Structured Output
JSON Output Files
Web Dashboard
Clinical Reports

Key Modules

- **Medical Image Analysis**: OpenCV preprocessing, CLIP classification, quality assessment
- **Clinical Text Processing**: BioBERT encoding, entity extraction, condition identification
- **Multimodal Integration**: LangChain orchestration, feature fusion, contextual analysis
- **Dual LLM Analysis**: Gemini primary assessment, Llama clinical verification

6. Results and Analysis

Validation Performance

- Success Rate: 100% (5/5 test cases)
- Average Processing Time: 10.1 seconds
- API Reliability: 100% uptime

Case Analysis Results

Test Case	Medical Condition	Detected Image Type	Clinical Correlation	Processing Time
1	Community-acquired Pneumonia	Chest X-ray	✅ Excellent	10.38s
2	Chronic Migraine	Brain MRI	✅ Strong	10.33s
3	Osteoarthritis	Knee X-ray	✅ High Confidence	9.63s
4	Post-appendectomy	Abdominal X-ray	✅ Consistent	8.87s
5	COVID-19 Pneumonia	Chest X-ray	✅ Strong	11.29s

General-Purpose Capabilities

- **Supported Inputs**: X-rays, CT, MRI, Ultrasound, DICOM + clinical texts
- **Key Features**: Real-time analysis, batch processing, adaptive handling
- **Architecture**: Modular design for unlimited case variations

7. Conclusion and Future Work

7.1 Summary of Contributions

- **Unified Multimodal Framework**: Integrated image analysis with clinical text processing
- **Dual LLM Validation**: Combined Gemini 2.0 and Llama 3.1 for clinical reliability
- **Medical-Optimized AI**: Adapted general models for healthcare applications
- **Structured Outputs**: Comprehensive JSON reports with clinical assessments

System Versatility

Validated with five test cases (100% success), the assistant processes:

- Diverse medical imaging modalities
- Various clinical documentation formats
- Real-time medical questions
- Batch patient case processing

7.2 Future Work

- Integration with PACS and EHR systems
- Enhanced domain-specific medical vision models
- Real-time processing for emergency scenarios
- Multi-language and specialty-specific modules
- Federated learning for privacy preservation

8. References

i. Google AI Gemini Models: https://ai.google.dev/

ii. Groq API: https://groq.com/

iii. Radford, A., et al. "CLIP." ICML 2021. DOI: 10.48550/arXiv.2103.00020

iv. Lee, J., et al. "BioBERT." Bioinformatics, 2020. DOI: 10.1093/bioinformatics/btz682

v. Wang, Z., et al. "MedCLIP." EMNLP 2022. DOI: 10.48550/arXiv.2210.10163

vi. LangChain: https://www.langchain.com/

Acknowledgement

I would like to express my sincere gratitude to the course instructors and the HAAI++ program for providing this opportunity to work on a challenging and meaningful capstone project. Special thanks to the open-source AI community for developing the tools and frameworks that made this project possible. The comprehensive documentation and support resources for Gemini API, Groq API, and the various AI libraries were invaluable throughout the development process. This project has significantly enhanced my understanding of multimodal AI systems and their practical applications in healthcare technology.