Agentic Multimodal Medical Assistant

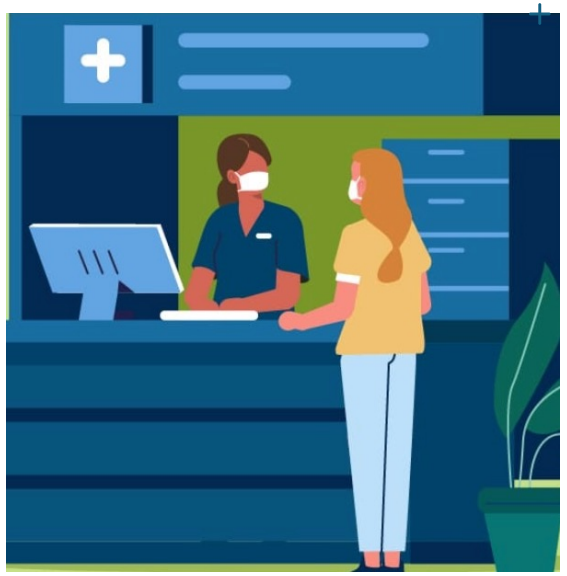
Using MultiCaRe Dataset

CARE CREW

Navya Kasula | Prabhu Kiran Gummadi | Gopikrishna Dengu | Shamitha Reddy Cheedu | Kruthi Reddy Kasarla

PROJECT OBJECTIVES

- Analyze multimodal medical data (images + captions + metadata)
- Build text and image embeddings
- Implement FAISS-based similarity search
- Generate LLM-based medical explanations
- Demonstrate an **agentic AI workflow** in a healthcare setting



DATASET OVERVIEW (MultiCaRe)

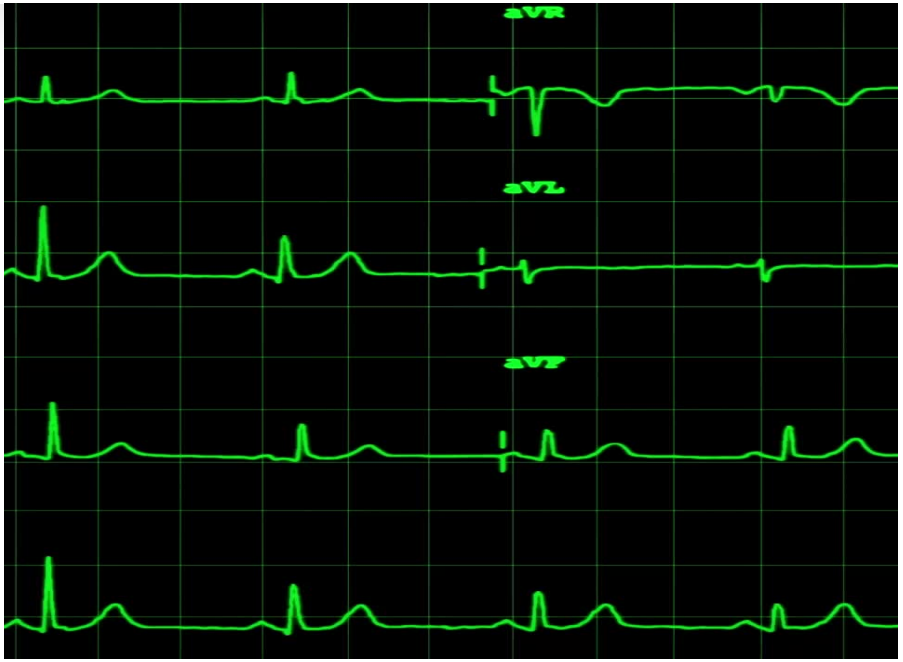
- Multimodal medical dataset including:

a. Radiology images (CT / MRI / X-ray)

b. Pathology images

c. Expert-written captions

- Metadata fields (image_type, generic_label, finding, laterality, site)
- ~135,000 caption records
- ~11,000 matched images
- ~5,000 cleaned samples used for analysis and modeling



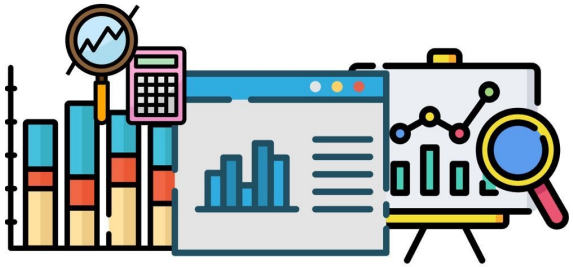
PROJECT STRUCTURE

Analytical understanding

- Cleaning & preprocessing
- EDA
- Feature engineering
- Logistic regression + RFE

Working multimodal model pipeline

- Text embeddings using transformers
- Image embeddings using ResNet
- FAISS retrieval
- LLM answer generation
- Agentic multimodal workflow demonstration



DATA CLEANING

- Standardized missing values
- Filled categorical NaN using mode
- Converted captions marked as empty to “Unknown”
- Matched images to caption entries using ID mapping
- Ensured complete multimodal pairs for modeling

Data Cleaning

```
captions_df['caption_clean'] = captions_df['caption'].apply(clean_caption)
captions_df.head()
```

	file_id	file	main_image	patient_id	license	caption	chunk	generic_label	pathology_test	image_type	image_te
0	file_000004	PMC10000323_jbsr-107-1-3012-g3_undivided_1_1.jpg	PMC10000323_01_jbsr-107-1-3012-g3.jpg	PMC10000323_01	CC BY	Pathological result.	[]	[]	[]	[]	
1	file_000005	PMC10000728_fmed-09-985235-g001_A_1_3.jpg	PMC10000728_01_fmed-09-985235-g001.jpg	PMC10000728_01	CC BY	Intraoperative exploration revealed a teratoma...	['teratoma', 'rectal', 'posterior', 'uterine w...	['Histology', 'Site', 'Position', 'Site', 'Lat...	[]	[]	
2	file_000006	PMC10000728_fmed-09-985235-g001_B_2_3.jpg	PMC10000728_01_fmed-09-985235-g001.jpg	PMC10000728_01	CC BY	The teratoma was disconnected from the posteri...	['teratoma', 'posterior', 'uterine wall', 'rig...	['Histology', 'Position', 'Site', 'Laterality'...	[]	[]	
3	file_000007	PMC10000728_fmed-09-985235-g001_C_3_3.jpg	PMC10000728_01_fmed-09-985235-g001.jpg	PMC10000728_01	CC BY	Gross observation of the specimen (C).	[]	[]	[]	[]	
4	file_000008	PMC10000728_fmed-09-985235-	PMC10000728_01_fmed-	PMC10000728_01	CC BY	The specimen contains hair.	['bone',	['Site', 'Histology']	[]	[]	

EDA SUMMARY

Caption Length Insights

- Mean ≈ 137 characters
- Long-tailed distribution
- Longer captions typically correspond to abnormal findings

Image Type Distribution

- CT, MRI, and X-ray most frequent
- Pathology images contain most detailed captions

METADATA PATTERNS

- Pathology → long, rich captions
- X-rays → short, structured captions
- Strong relationships observed between:
- **generic_label** and **finding**
- **image_type** and **caption length**

FEATURE ENGINEERING

- Created:
- **caption_length** (numeric)
- **has_finding** (binary)
- **generic_label_enc**
- **image_type_enc**
- Missing values handled → complete dataset

LOGISTIC REGRESSION + RFE RESULTS

- **Top Predictors of Findings:**

1. **caption_length**
2. **generic_label_enc**
3. **image_type_enc**

- Metadata + text structure alone provide consistent signals about abnormalities.

MULTIMODAL MODELING OVERVIEW

Text Model

- SentenceTransformer (all-MiniLM-L6-v2)
- Converts captions → 384-dimensional embeddings

Image Model

- ResNet-50 pretrained
- Converts images → 2048-dimensional feature vectors

Storage

- Two FAISS indexes (text + image) for retrieval

TEXT RETRIEVAL PIPELINE

Query → Embedding → FAISS Retrieval → LLM Explanation

System retrieves semantically similar captions
Example:
“Left lung opacity” → returns clinically relevant captions with similar findings

IMAGE RETRIEVAL PIPELINE

Image → ResNet Embedding → Image FAISS Search → Similar Images

- Used for visual comparison:
- MRI → similar MRI scans
- Chest CT → similar CT cases

LLM RESPONSE GENERATION

LLM takes:

- Top-k retrieved captions
- Optional image context

Outputs:

- Clinical explanation
- Summary of findings
- Possible interpretations

This forms a **retrieval-augmented medical assistant**.

AGENTIC WORKFLOW

Steps:

1. User inputs text/image
2. System retrieves top relevant cases
3. Constructs contextual prompt
4. LLM generates medical explanation
5. (Optional) Reflection/correction

Mimics a real-world medical AI assistant with reasoning capability.

SYSTEM DEMONSTRATION

Text query → similar caption retrieval output

```
response = chat_with_model("Hi, I have a headache. What should I do?")
print(response)
```

```
system
You are a helpful health assistant. Do not give medical diagnosis. Provide general guidance only.
user
Hi, I have a headache. What should I do?

If you're experiencing a headache, there are several steps you can take to alleviate the discomfort and promote healing:

1. Rest: The first step is to rest your head as much as possible. This can help reduce inflammation in the brain and muscles, which can contribute t

2. Apply heat or cold: Applying a warm compress or ice pack to your forehead, temples, neck, or back can help relieve pain and numbness. If you pref

3. Over-the-counter pain relievers: Over-the-counter pain relievers such as ibuprofen (Advil, Motrin) or acetaminophen (Tylenol) can help reduce inf

4. Relaxation techniques: Deep breathing exercises, meditation, or progressive muscle relaxation can help reduce stress and tension in your head and

5. Avoid triggers: Identify any triggers that may cause your headaches and avoid them as much as possible. These could include certain foods, drinks
```

```
response = chat_with_model("Hi, I have a fever. What should I do?")
print(response)
```

```
system
You are a helpful health assistant. Do not give medical diagnosis. Provide general guidance only.
user
Hi, I have a fever. What should I do?

If you have a fever, there are several things you can do to help alleviate your symptoms and speed up the recovery process:

1. Rest: The most important thing is to get plenty of rest. Your body needs time to fight off the infection and recover from the fever.

2. Hydration: Drink plenty of fluids, such as water, clear broths, or electrolyte-rich drinks like coconut water. This will help replace fluids lost

3. Take over-the-counter medication: If your fever is accompanied by other symptoms like cough, headache, or body aches, you may need to take an ove
```

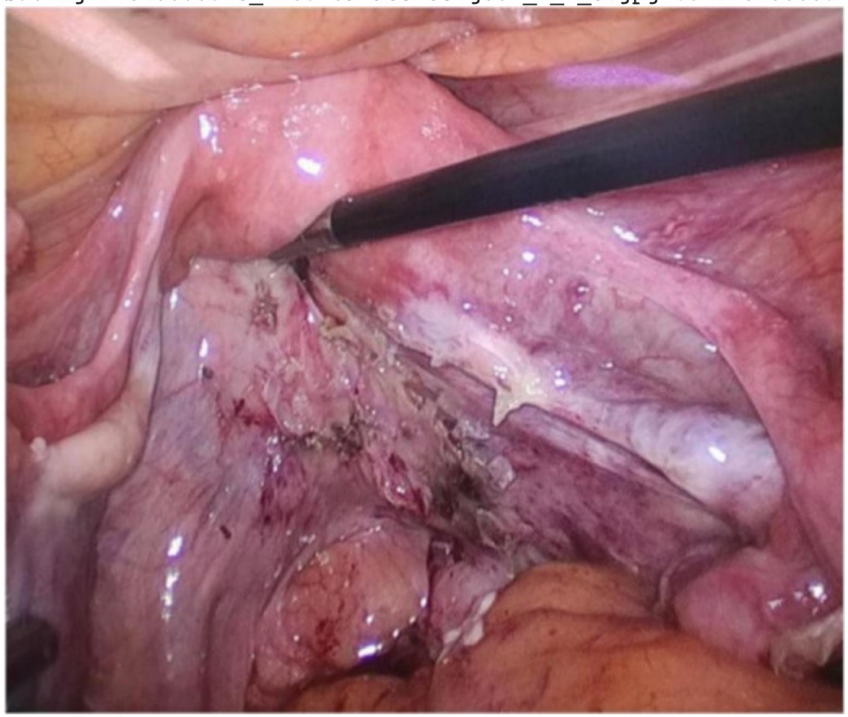
SYSTEM DEMONSTRATION

- Image query → similar image clustering
- LLM → structured diagnostic explanation

```
print("Description:", response)
```

Choose Files no files selected Upload widget is only available when the cell has be

```
Saving PMC10000728_fmed-09-985235-g002_B_2_3.jpg to PMC10000728_fme
```



```
Description:
user

Describe this image.

assistant
The image depicts a surgical procedure being performed through an o
```

KEY OUTCOMES

- Working multimodal prototype
- Retrieval-enabled similarity engine
- Metadata strongly correlates with findings
- LLM explanations contextual and medically coherent
- Demonstrates full agentic pipeline from input → retrieval → reasoning → output

LIMITATIONS

- No end-to-end multimodal fusion model
- No medical-domain fine-tuning (BioCLIP/LLaVA-Med)
- Sparse metadata in some fields
- Limited compute for large-scale experiments
- No clinical validation

FUTURE WORK

Federated Learning

- Train across hospitals without sharing raw data
- Enhances privacy & increases real-world potential

Fine-tuned CLIP-Style Medical Model

- Joint image–text training
- Improved alignment & accuracy

CONCLUSION

- Successful demonstration of a multimodal agentic assistant
- Found significant patterns in medical text + image metadata
- Functional retrieval + LLM reasoning pipeline
- Strong foundation for future clinical AI systems

THANK YOU