OAAgent: A Multimodal LLM Agent Clinical Assistant for Precision Osteoarthritis Care

Pegah Ahadian and Qiang Guan (Kent State University, OH)

Motivation

Knee OA affects 300M+ people worldwide. Current predictive tools are limited:

- Rely on single data sources (usually X-rays).
- Function as black boxes with no explanations.
- Poor integration into clinical workflows.
- Medical AI fails not because of accuracy, but because it cannot explain its reasoning in clinically meaningful terms.

Solution:

- Fusion Transformer for multimodal integration
- C-TRAG temporal retriever for cross-visit trajectory alignment
- Chain-of-Thought reasoning and clinical note summarization for transparent recommendations

Research Questions

How do I create this AI framework that reasons like clinicians while maintaining computational efficiency?
Three Design Principles I Tested:

Reasoning Decomposition: Medical decisions follow sequential logical steps
Multimodal Integration: Clinicians don't "fuse features" - they narrate observations.
Dynamic Adaptability: Real clinical reasoning adapts to incomplete information

Multimodal Fusion

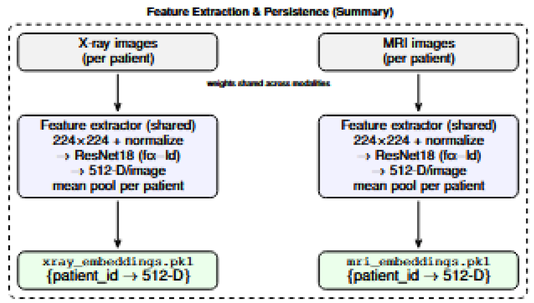
Inputs: X-ray, MRI, and longitudinal clinical variables.

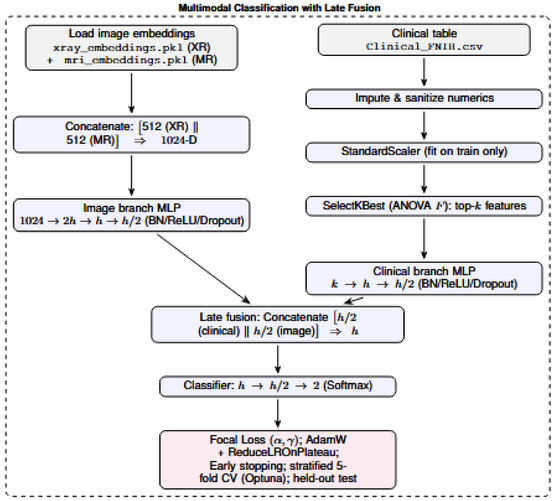
Dataset: Validated on OAI FNIH cohort for OA progression prediction.

Fusion Transformer: Attention-based cross-modal architecture that learns complementary signals from imaging and structured clinical data.

Outcome: A unified patient trajectory representation that improves progression prediction compared to monomodal models.

- Improved OA progression prediction accuracy
- Inner Cross-validation Mean Validation Accuracy: 0.7941 ± 0.0356
  Model Test Accuracy (on Held-Out Test Set): 0.7536





Temporal Retrieval (C-TRAG)

Capturing Similar Trajectories in Osteoarthritis Progression. The system learns which perspective best explains similarity for a given query.

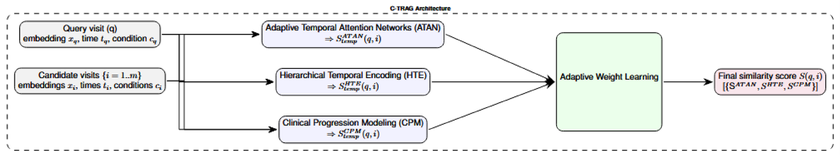
Mechanism: In C-TRAG, the similarity between the query visit q and each candidate visit i is not just one measure (like cosine similarity). Instead, we designed three different ways of measuring similarity, and then combined them adaptively

- **Adaptive Temporal Attention Networks:** Model the importance of when visits occur, which time visit is more informative.
- **Hierarchical Temporal Encoding:** Captures both short-term and long-term dependencies, captures sudden vs gradual progression.
- **Clinical Progression Modeling:** Injects clinical priors, captures clinically progression drivers.
- **Adaptive Weight Learning:** Instead of choosing one similarity function, it learns the weights.

Dataset: Developed and tested on MIMIC-III/IV (clinical trajectories) and MIMIC-CXR.

- Identify high-risk patients for preventive care
- Prioritize care for patients most likely to progress
- Supports reasoning over disease evolution rather than isolated snapshots.

** Combining C-TRAG with OAAgent for full multimodal reasoning is in progress. Currently under IRB review for access to OA-related physician notes.



$$S(q,i) = \alpha S^{\text{ATAN}}(q,i) + \beta S^{\text{HTE}}(q,i) + \gamma S^{\text{CPM}}(q,i)$$

$$z = [\mathbf{x}_q \,\|\, h^{\text{local}}(q) \,\|\, h^{\text{global}}(q) \,\|\, \mathbf{c}_q]$$

$$[a,b,c] = \mathbf{MLP}_{\text{gate}}(z)$$

$$[\alpha,\beta,\gamma] = \mathbf{softmax}([a,b,c])$$

Trustworthy AI

Lack of transparency in AI systems undermines clinical trust and usability, especially for progression prediction. We proposed **Transparent Dynamic Environment** that Prompt as Cognitive Architecture: prompt + patient data → reasoning cycle → explained prediction

Chain-of-Thought (CoT) Reasoning: Produces step-by-step clinical rationale behind each prediction.

Transparency: Clinician-facing explanations link recommendations back to input data and retrieved patient cases.

- Transparent LLM-based explanations
- Alignment with Trustworthy AI principles

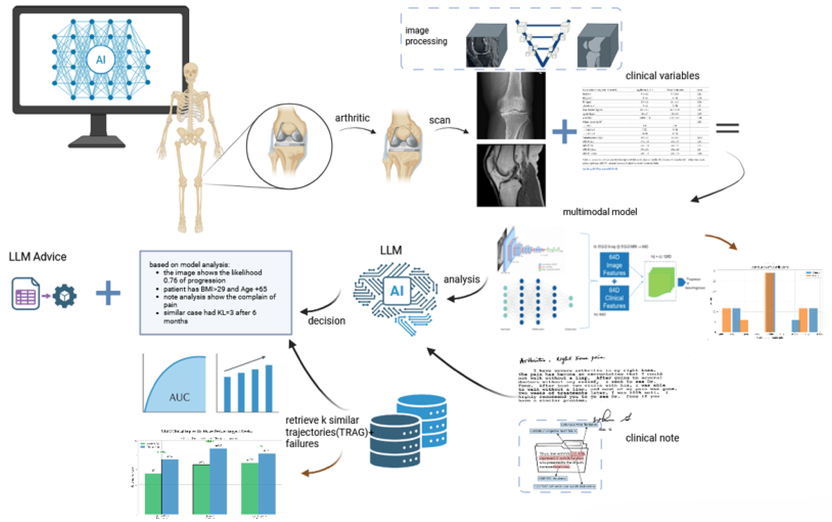
```
Run the environment
Diagnose patient 9026416

                    Generating Agent response...
                    Thought 1: Retrieve the patient's clinical data and imaging.
                    Action 1: retrieve[9026416]
                    Observation 1: Retrieved clinical data: [0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 5.8, 5.293, 4.737, 32.6, 64.0, 3.0, 1.0, 0.0, 1.0625], image data shape: (1024,)
Process the data with the multimodal model.
                    Thought 2: Process the data with the multimodal model.
                    Action 2: process_multimodal[clinical_data|image_data]
                    Observation 2: The likelihood of progressing is 0.5279
                    Thought 3: Probability 0.5279 > 0.5, so this is a PROGRESSOR.
                    Action 3: finish[progressor]
                    Observation 3: Episode finished
Final answer!
                    Final Info: {'steps': 3, 'answer': 'progressor'}
                    Final Answer: progressor
What are the important features in this prediction?
                    P01KPMEDCV, V06XRJSL, P01BMI, P02RACE, V06MCMJSW
What is the contribution of modality?
                    Image: %2.1, Clinical: %97.9
```

OAAgent Decision-Making Process

Workflow

OAAgent is an AI-powered assistant that integrates multimodal data, reasons like a clinician, and delivers transparent outputs.

- **Multimodal integration:** OAAgent unifies heterogeneous inputs — radiology images (X-ray, MRI), longitudinal clinical variables, and physician notes — into a single coherent representation.
- **Clinical reasoning:** The agent leverages temporal retrieval (C-TRAG), fusion transformers, and reinforcement learning with Chain-of-Thought to reason like a clinician across visits and trajectories.
- **Transparent outputs:** Every recommendation is evidence-linked and interpretable, providing saliency maps, feature attributions, retrieved analog cases, and cited note spans to ensure traceability and clinician trust.



OAAgent Workflow

Conclusion

- OAAgent unifies X-ray/MRI, longitudinal clinical variables, and physician notes with a fusion transformer, the C-TRAG temporal retriever, and an agentic decision layer.
- Delivers stronger progression prediction, better temporal calibration, and case-based explanations than unimodal and standard multimodal baselines.
- Explanations are evidence-linked (saliency maps, feature attributions, cited note spans, retrieved analog cases), supporting clinician trust.

By C-TRAG we addressed crucial questions in real world OA dignosis:

- Are these patients similar when viewed at the right time in disease progression?
- Are their overall trajectories, short-term and long-term, aligned?
- Do they share the same clinical drivers (BML, JSL, meds, BMI)

Acknowledgement: Collaborative work with **Cleveland Clinic** clinical expertise, data access, and iterative validation.

Implications

- **Enhanced Clinical Support:** Earlier, individualized risk stratification with *evidence-linked* explanations (saliency, attributions, cited note spans, analog cases).
- **C-TRAG Retriever:** Semantically *and* temporally aligned case retrieval for trajectory-aware counseling and shared decision-making.
- **LLM–Healthcare Integration:** MCP-based EHR/PACS interoperability; Extract→Abstract notes; clinician-editable summaries embedded in workflow.
- **Clinical Partnership:** Co-developed and validated with Cleveland Clinic on OAI/FNIH cohorts; generalization checks with MIMIC text.

Next:

- scale MRI cohorts
- deploy MCP radiomics services
- multi cross dataset validation
- refine RL policy with off-policy & causal evaluation
- assess health economic impact.

Artifacts & Links

Demo: `https://drive.google.com/file/d/10j6FWWhtNnUTqNb13oXUkzLNBCFDg-kW/view?usp=sharing`
Code/Docs: `https://github.com/pegahahadian/Doctoral-Showcase-SC25/tree/main`

Pahadian@kent.edu Google Scholar LinkedIn