One Agent, Many Models: An AI-Agent Approach to Universal Tumor Segmentation

Jiaqi Liu[1,2], Dan Nguyen, PhD[1,2], Steve B. Jiang, PhD[1,2] and Ti Bai, PhD[1,2]
[1]Medical Artificial Intelligence and Automation (MAIA) Lab, Department of Radiation Oncology, UT Southwestern Medical Center, Dallas, TX
[2]Department of Radiation Oncology, UT Southwestern Medical Center, Dallas, TX
Contact: jiaqi.liu@utsouthwestern.edu

MEDICAL PHYSICS COMMUNITY · INSPIRATION · INNOVATION · IMPACT
2026 JULY 19–22 VANCOUVER, BC
JOINT AAPM | COMP MEETING

MAIA Laboratory
UT Southwestern Medical Center
Radiation Oncology

Background

Recent advances in deep learning have produced numerous high-performing tumor segmentation models across different imaging modalities and anatomical sites.[1] However, despite their accuracy, many models remain difficult to translate into clinical practice because selecting appropriate models, preprocessing input data, deploying models from open-source codes, and output QA require substantial computational expertise.[2]

This work proposes an agentic AI framework that bridges the gap between state-of-the-art segmentation models and real-world clinical deployment. Instead of developing a single universal segmentation model, the framework leverages large language model (LLM)-driven agents to intelligently select and orchestrate existing task-specific models, enabling clinicians to access an expandable ecosystem of segmentation tools through an automated and user-friendly workflow (Figure 1).

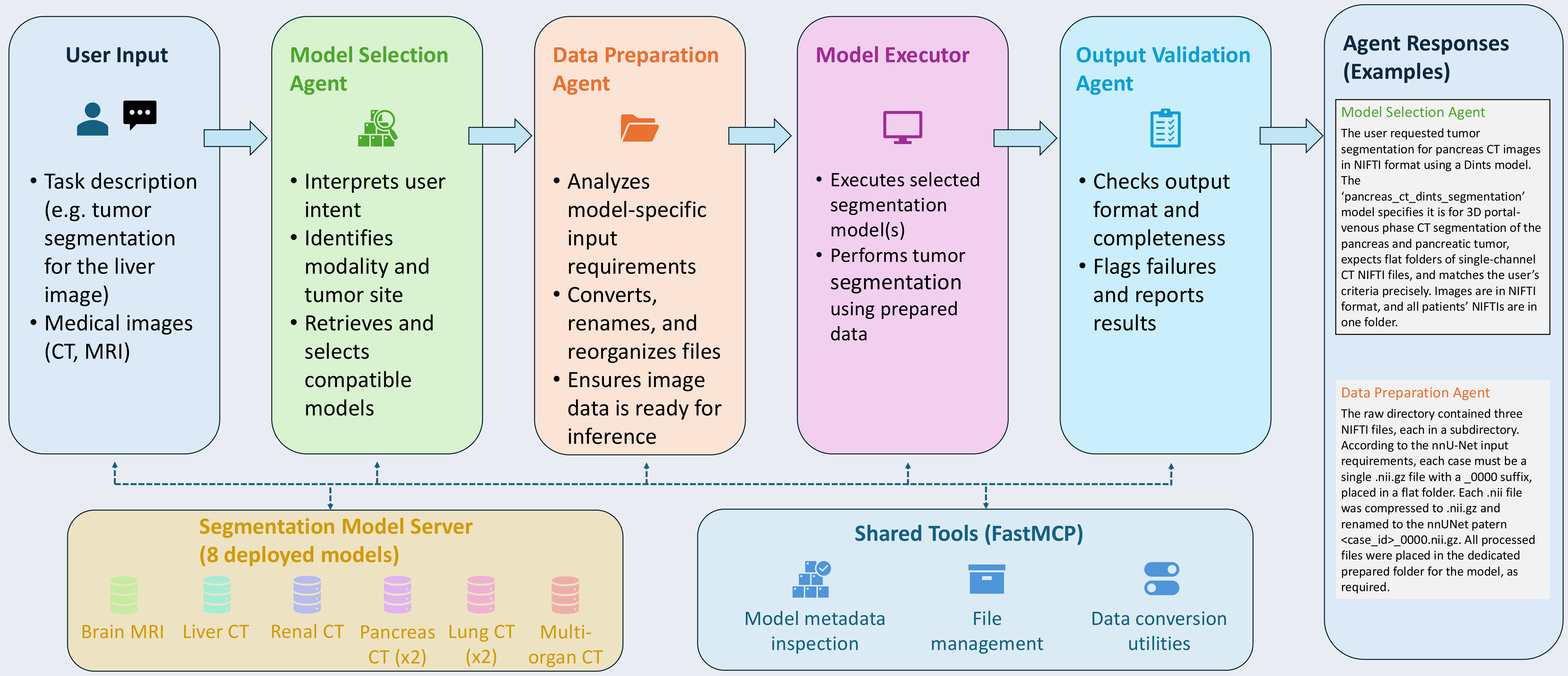

Figure 1. Overview of the tumor segmentation agentic AI framework

METHODS AND MATERIALS

The proposed **sequential reasoning** framework decomposes the tumor segmentation workflow into four stages:

1. **Model Selection Agent**: interprets user requests, identifies imaging modality and tumor site, and selects compatible segmentation models based on model documentation.
2. **Data Preparation Agent**: analyzes model-specific input requirements and automatically reformats, renames, and organizes medical images.
3. **Model Executor**: performs segmentation inference using Python-based execution pipelines.
4. **Output Validation Agent**: evaluates segmentation completeness, format compatibility, and identifies potential execution failures.

Agents interacted with the execution environment through Python functions registered as FastMCP tools, allowing automated file management, metadata inspection, and model execution. Agents were prompted with structured role descriptions and operational rules.

Eight publicly available tumor segmentation models were deployed on a local server, covering brain MRI, renal CT, liver CT, pancreas CT, lung CT, and multi-organ CT segmentation tasks.

The system was evaluated using 48 cases featuring 6 testing scenarios designed to simulate realistic clinical challenges, including ambiguous user input, inconsistent data organization, model malfunction, and worst-case stress conditions. A case was counted successful if the system selected a compatible model, completed inference, and performed automated checks.

RESULTS

Overall Performance

- 77.1% end-to-end success rate
- 0.85 end-to-end F1 score
- 87.5% best-case success rate
- 62.5% worst-case success rate

Key Findings

- Individual agents achieved high reliability (F1: 0.91–1.00)
- Framework successfully handled:
  - heterogeneous input organization
  - model-specific requirements
  - execution failures

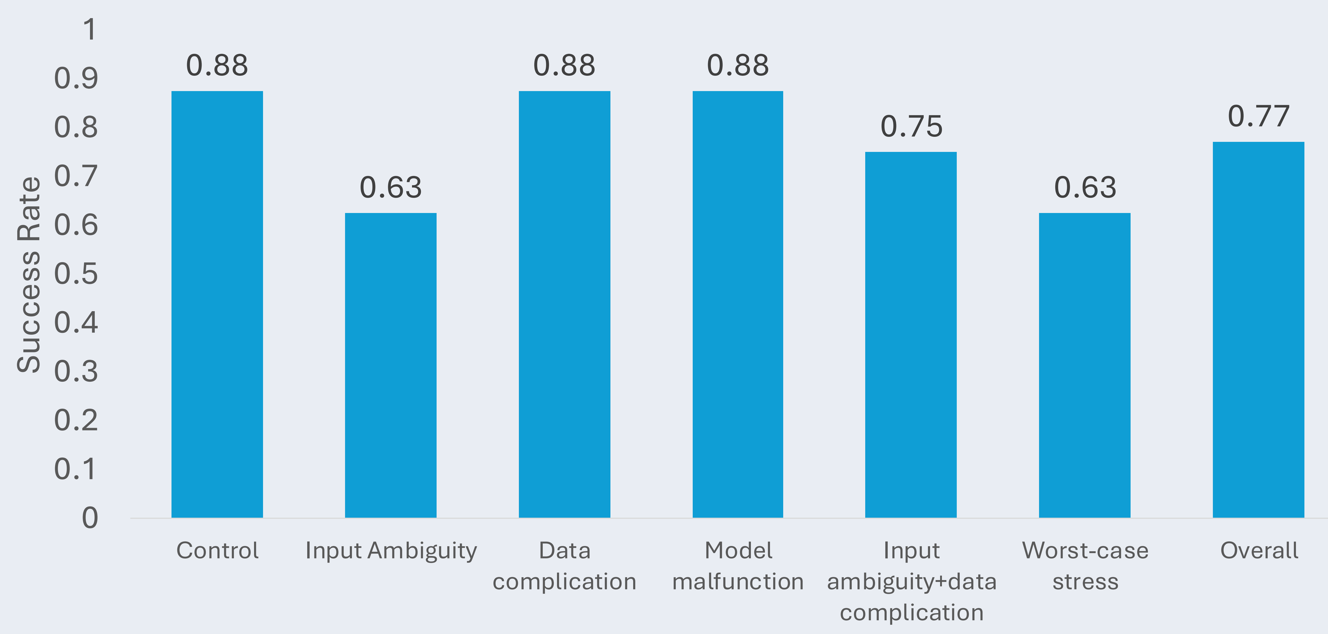

Figure 2. The agent's end-to-end success rate under different test scenarios

Agent	Precision	Recall	F1 Score
Model Selection	0.83	1	0.91
Data Preparation	0.94	1	0.91
Output Validation	1	1	1
End-to-end	**0.74**	**1**	**0.85**

Table 1. Sub-agents' performance across all test cases

CONCLUSIONS

An LLM-driven agentic AI framework was developed to automate tumor segmentation model selection, data preparation, inference, and output verification. The system successfully integrates diverse segmentation models into a unified workflow and demonstrates robustness when facing real-world data heterogeneity and execution complications.

This approach provides a scalable and accessible pathway for translating rapidly expanding AI segmentation research into clinical workflows. Future development will extend the framework toward interactive end-to-end assistance for tumor diagnosis, treatment planning, and response assessment.

DISCUSSION

Key Advantages

The results highlight the advantages of our modular agent-driven framework:

- Enables integration of specialized segmentation models instead of relying on a single universal foundation model
- Reduces domain-shift limitations through task-specific model selection[3]
- Improves accessibility by automating computationally demanding steps
- Provides transparency by separating user intent interpretation, data preparation, inference, and quality assurance into independent reasoning steps
- Provides an expandable framework for continuously incorporating new AI models

Limitations and Future Direction

User input ambiguity—such as missing modality or tumor site information—remains a challenge. Interactive agent–user communication will be incorporated to improve handling of incomplete or ambiguous clinical requests.

References

1. Jiang H, Diao Z, Yao YD. Deep learning techniques for tumor segmentation: a review. J Supercomput. 2022;78(2):1807-1851. doi:10.1007/s11227-021-03901-6
2. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med. 2019;17(1):195. doi:10.1186/s12916-019-1426-2
3. He S, Bao R, Li J, et al. Computer-Vision Benchmark Segment-Anything Model (SAM) in Medical Images: Accuracy in 12 Datasets. arXiv. Preprint posted online May 5, 2023:arXiv:2304.09324. doi:10.48550/arXiv.2304.09324