# Large language model-based organ contouring/segmentation for cancer in radiotherapy

Integrating vision and language models for enhanced medical analysis.


Yashaswi Patki
Tanay Gawade

## Challenges in Medical Image Segmentation:

- Complex 3D medical image analysis
- Limited context understanding
- Need for multimodal integration
- High accuracy requirements for medical applications

## Current Limitations:

- Traditional methods lack contextual understanding
- Single-modal approaches miss valuable information
- Limited integration of medical reports

# Technical Stack

Core Technologies:

- Deep Learning Framework: PyTorch 2.0.1
- Primary Libraries:
  - MONAI (Medical Open Network for AI)
  - Hugging Face Transformers
  - Llama-2 (Meta AI)
  - TensorBoardX

System Requirements:

- GPU: NVIDIA GPU (45GB+ VRAM)
- RAM: 64GB+
- Storage: 100GB+
- OS: Linux (Ubuntu) / Windows with WSL

# Project Overview

## Core Components

2. Language Understanding

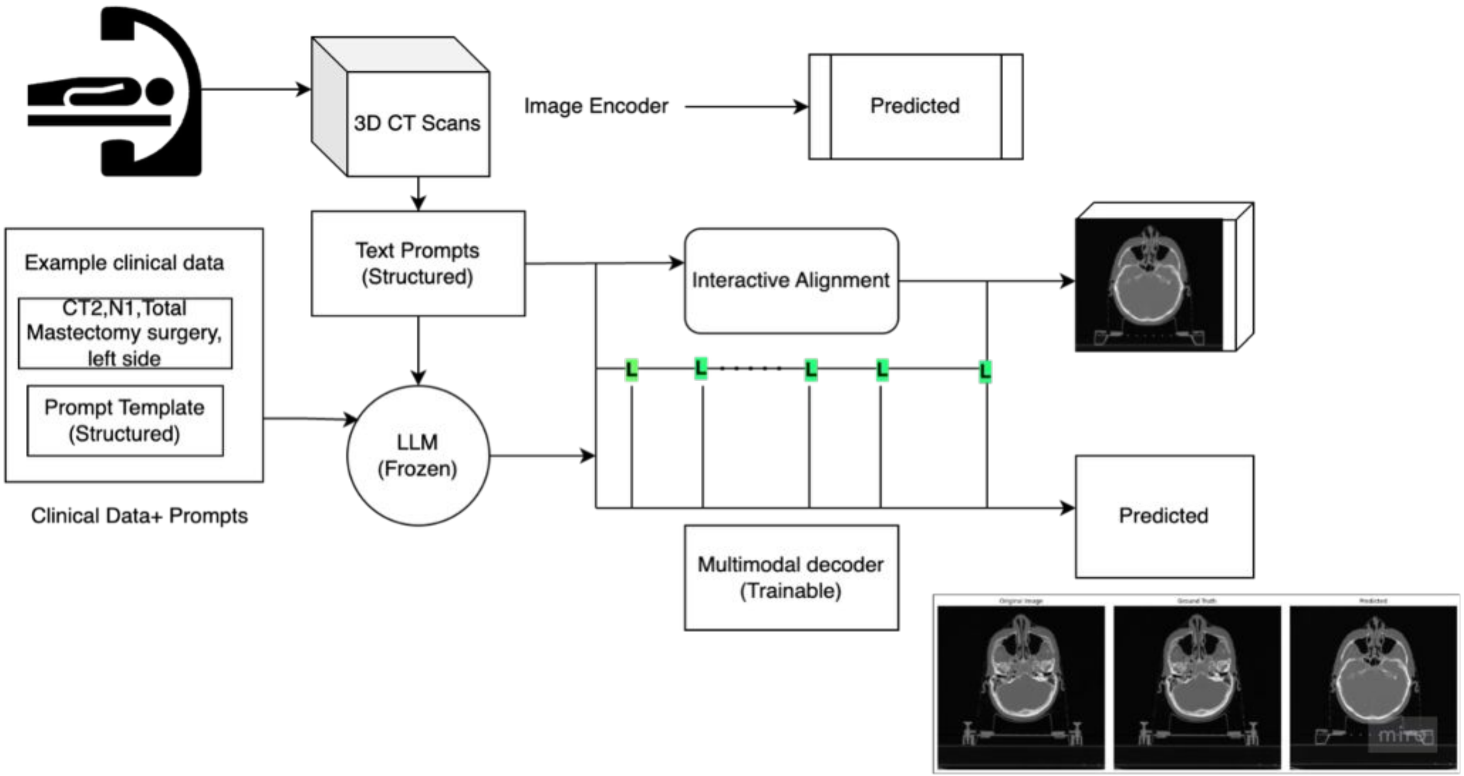
- ➔ Llama-2 7B model integration
- ➔ Medical report analysis
- ➔ Context-aware processing

1. Image Processing Pipeline

- ➔ 3D UNet-based architecture
- ➔ Multi-scale feature extraction
- ➔ Region of Interest: 64×352×352

3. Multimodal Fusion

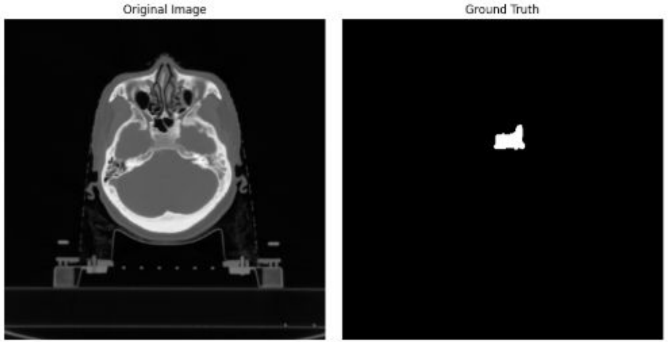
- ➔ Custom attention mechanism
- ➔ Cross-modal feature integration
- ➔ Enhanced segmentation accuracy


3D CT Scans
Image Encoder
Predicted
Example clinical data
CT2,N1,Total Mastectomy surgery, left side
Prompt Template (Structured)
Clinical Data+ Prompts
Text Prompts (Structured)
Interactive Alignment
L
L
L
L
L
LLM (Frozen)
Predicted
Multimodal decoder (Trainable)

# Imaging and Textual Data Processing

➔ Imaging Data:
- Voxel standardization
- Intensity normalization
- Spatial cropping around tumor regions

➔ Textual Data:
- De-identification of patient information
- Tokenization using LLaMA-2 tokenizer
- Structured prompts for accurate context embedding




| Number | Unit No | Subsite 1 | Pathology | icT | icN | icM | pT | pN | pM | Stage | Chemotherapy | RT aim | Remark |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1573 | Head | SCC | T2 | N0 | M0 | | | | icT2N0M0 | Adjuvant | Definitive | Definitive RT (BCS) |
| 2 | 1612 | Head | SCC | T1c | N1 | M0 | | | | pT1cN1M0 | Adjuvant | Definitive | Definitive RT (BCS) |
| 3 | 1672 | Head | SCC | T1 | N0 | M1 | | | | pT1N0M0 | Adjuvant | Definitive | Definitive RT (BCS) |
| 4 | 1688 | Head | SCC | T2 | N1 | M2 | | | | pT2N1M0 | Adjuvant | Definitive | Definitive RT (BCS) |

# Training Strategy

- ➔ Pretraining: Modality alignment with contrastive learning
- ➔ Fine-tuning: Supervised segmentation with Dice and Focal losses
- ➔ Optimization: AdamW optimizer, cosine scheduler, warmup
- ➔ Visuals: Loss curves and optimization graphs

# Results

## Dataset and Evaluation Metrics

Dataset: 100 annotated HNC patients, 3D CT volumes, clinical notes

Testing for 2 cases

| no | ID | Dice | IoU | HD |
|---|---|---|---|---|
| 0 | 1573 | 0.802 | 0.670 | 27.783 |
| 1 | 1571 | 0.804 | 0.672 | 2.209 |
| Avr | - | 0.803 | 0.671 | 14.996 |
| Std | - | 0.001 | 0.001 | 12.787 |

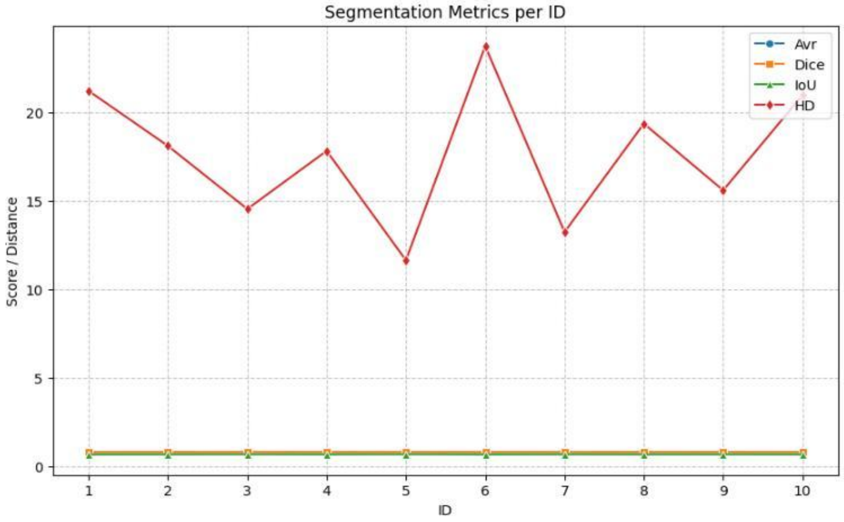
Testing for 5 cases

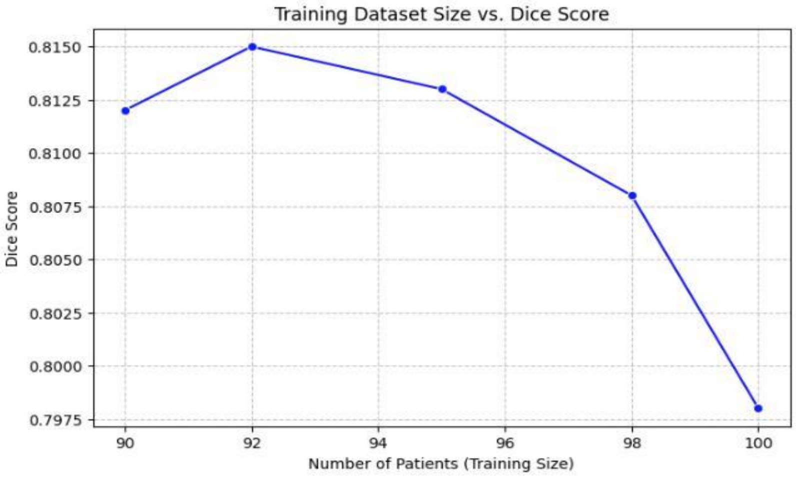
| no | ID | Dice | IoU | HD |
|---|---|---|---|---|
| 0 | 1573 | 0.802 | 0.670 | 5.569 |
| 1 | 1571 | 0.804 | 0.672 | 2.209 |
| 2 | 1543 | 0.810 | 0.672 | 7.934 |
| 3 | 1545 | 0.798 | 0.664 | 4.449 |
| 4 | 1573 | 0.807 | 0.670 | 10.573 |
| Avr | - | 0.802 | 0.666 | 7.546 |
| Std | - | 0.006 | 0.004 | 2.261 |

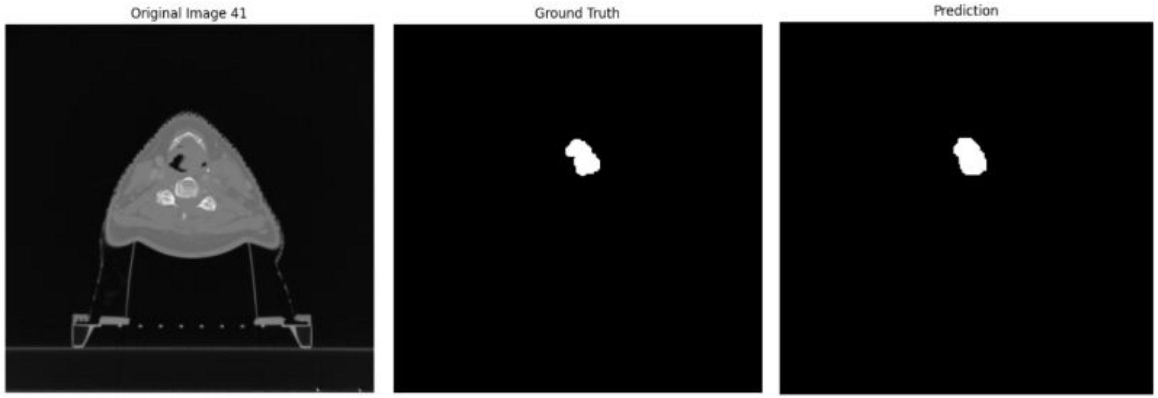
Testing for 10 cases

result_external1_report

| no | ID | Avr | Dice | IoU | HD | Report |
|---|---|---|---|---|---|---|
| 0 | 1 | 0.795 | 0.795 | 0.661 | 21.235 | |
| 1 | 2 | 0.800 | 0.800 | 0.667 | 18.121 | |
| 2 | 3 | 0.803 | 0.803 | 0.670 | 14.553 | |
| 3 | 4 | 0.798 | 0.798 | 0.665 | 17.843 | |
| 4 | 5 | 0.805 | 0.805 | 0.673 | 11.654 | |
| 5 | 6 | 0.796 | 0.796 | 0.662 | 23.776 | |
| 6 | 7 | 0.802 | 0.802 | 0.669 | 13.249 | |
| 7 | 8 | 0.799 | 0.799 | 0.666 | 19.384 | |
| 8 | 9 | 0.801 | 0.801 | 0.668 | 15.612 | |


Segmentation Metrics per ID
Avr
Dice
IoU
HD
Score / Distance
ID
0
5
10
15
20
1
2
3
4
5
6
7
8
9
10


Training Dataset Size vs. Dice Score
Dice Score
Number of Patients (Training Size)
0.8150
0.8125
0.8100
0.8075
0.8050
0.8025
0.8000
0.7975
90
92
94
96
98
100


Original Image 41
Ground Truth
Prediction

# Data Efficiency and Robustness

- **Data Efficiency Analysis:**
  - Maintained 83% performance with 90% of the data
  - Effective even in low-data scenarios
- **Robustness Analysis:**
  - Handles anatomical variability
  - Adapts to complex structures

# Clinical Impact

- ➔ Reduced Time for Manual Contouring
- ➔ Improved Treatment Planning and Precision
- ➔ Better Tumor Localization

# Future Work and Improvements

- ➔ Prompt Standardization and Fine-Tuning
- ➔ Multi-Institutional Validation
- ➔ Extension to MRI and PET Modalities
- ➔ Real-Time Deployment for Clinical Use

# THANK YOU !