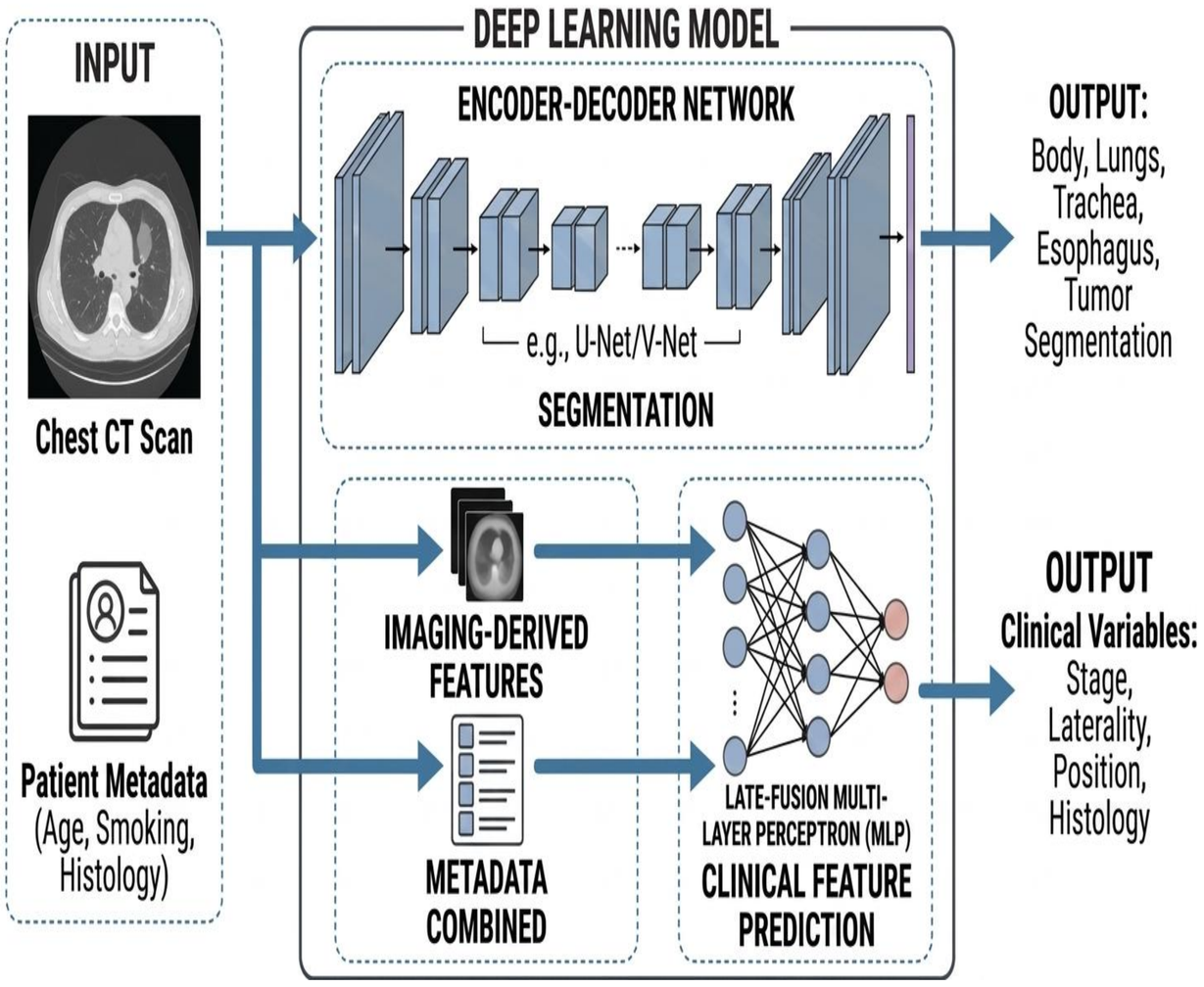
Multimodal AI for Joint Lung Tumor Segmentation and Clinical Feature Prediction from CT Scans

Mahmud Hasan Mostofa Kamal[1], Md. Nazrul Islam[1], Mohammad Monir Uddin[2]
[1]Department of Radiology and Imaging, Bangladesh Medical University, Dhaka, Bangladesh
[2]Department of Mathematics & Physics, North South University, Dhaka, Bangladesh

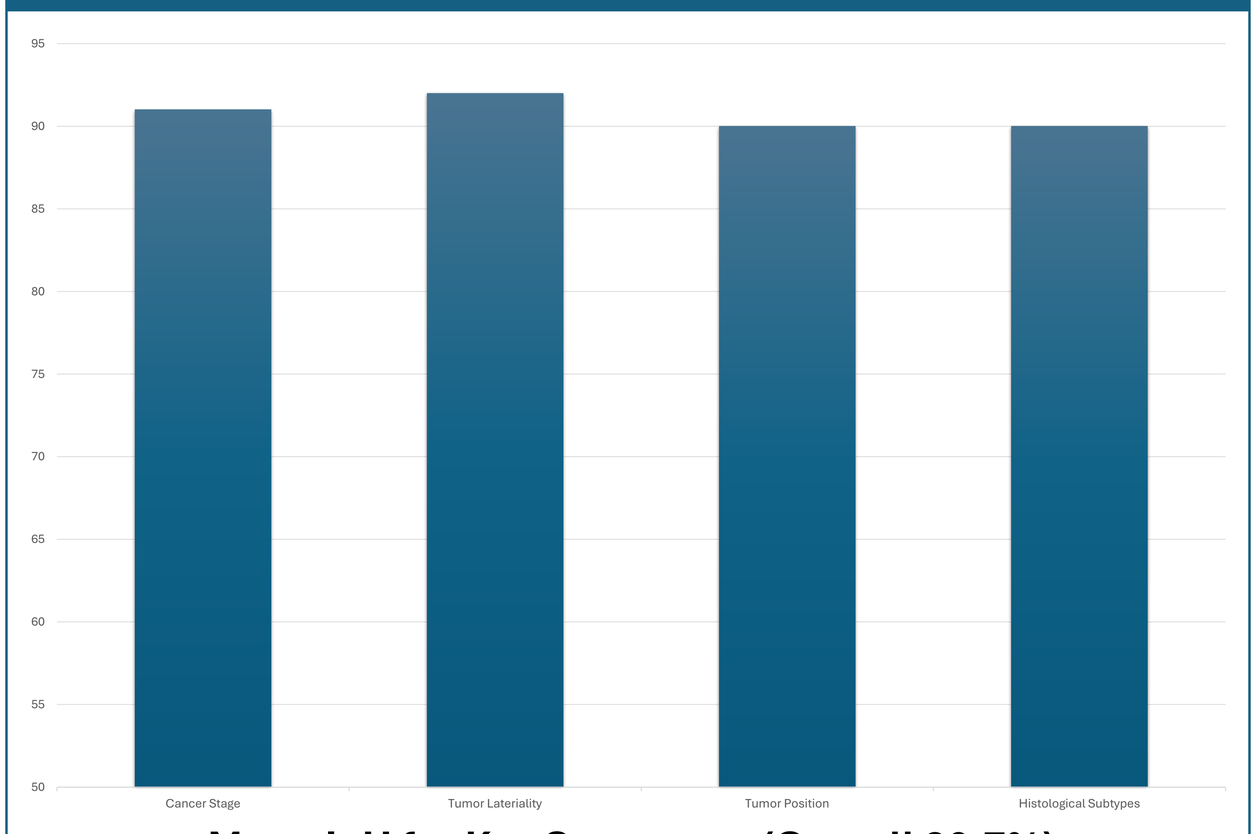
INTRODUCTION & OBJECTIVES

This multimodal deep learning model streamlines lung cancer analysis by merging high-resolution anatomical segmentation with clinical variable prediction in a single end-to-end system. By fusing CT imaging features with patient metadata, the model delivers a 90.8% mean IoU for segmentation and over 90% accuracy for predicting cancer stages and histology on diverse clinical data from Bangladesh. Integrated into a real-time graphical interface, it provides a powerful diagnostic "second opinion" designed to boost accuracy and workflow efficiency in both high-volume and resource-limited medical settings.

METHODOLOGY: MULTIMODAL ARCHITECTURE



SEGMENTATION PERFORMANCE



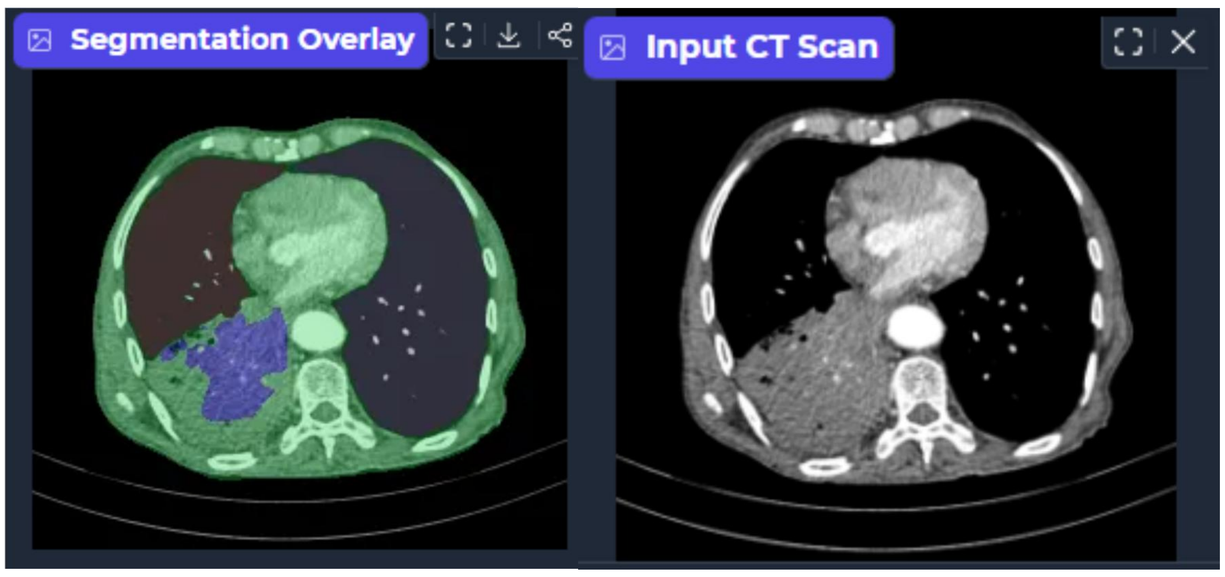
Mean IoU for Key Structures (Overall 90.7%)

QUANTITATIVE COMPARISON OF PROPOSED MULTIMODAL MODEL VS. BASELINE

Metric	Proposed Multimodal Model	Baseline (Imaging Only)
Segmentation IoU	90.8%	86.1%
Cancer Stage Accuracy	91.3%	78.4%
Tumor Laterality Accuracy	92.1%	80.2%
Tumor Position Accuracy	90.5%	75.1%
Histological Subtype Accuracy	90.8%	73.6%
Balanced Accuracy (mean)	91.2%	76.8%
Precision	90.4%	74.9%
Recall	91.7%	76.3%
F1 Score	91.0%	75.6%
Inference Time (per scan)	3.2s	2.8s

Performance comparison between the proposed multimodal model and imaging-only baseline across segmentation and clinical prediction tasks. All metrics are reported per CT scan on a single GPU.

GUI & CLINICAL DEPLOYMENT



Impression: Bilateral lung fields are clearly segmented with predominant right lung area compared to left. No visible trachea or esophageal segmentation in this slice. Presence of a measurable tumor region (~10.5% area), requiring further clinical correlation for characterization, staging, and localization.

CONCLUSION

This research successfully demonstrates the efficacy of an end-to-end multimodal deep learning model for automated lung cancer analysis. By integrating anatomical segmentation with clinical feature detection, the system provides a robust solution for enhancing diagnostic workflows, particularly in resource-constrained environments.