



Longitudinal Tracking of Meningioma Volume Using Deep Learning with Automated Confidence Measures

Pierre Nedelec, MS, MTM, UCSF - Rauschecker Sugrue Lab; Evan S. Bloch; Yassine Guennoun;
Mark McArthur, MD; Kevin Leu, MD; Jinchi Wei, MS; Radhika Bhalerao, MS; Shan McBurney-Lin, MBA, MD;
Leo P. Sugrue, MD, PhD; Andreas M. Rauschecker, MD, PhD

Introduction/Background

Meningiomas are the most common primary intracranial tumors and are frequently monitored with serial MRI. Clinical interpretation typically compares the current exam to a single prior, based on 2D diameter measurements, which can miss slow or multifocal growth. Volumetric assessment across multiple timepoints offers greater sensitivity but is too labor-intensive for routine clinical use. There is a critical need for automated, reproducible tools that enable longitudinal volumetric tracking of meningiomas over time and enhance clinical decision-making.

Methods/Intervention

We developed a pipeline combining 3D segmentation models with custom post-processing methods to track meningioma volumes across serial MRI exams. The models were trained on T1-weighted post-contrast sequences and segmentations from 1655 exams from 788 unique patients. From 38,000 meningioma exams that exist at UCSF, we filtered for patients with a single meningioma, no other tumor, and extracted resection status, leveraging a secure PHI-compliant version of GPT-4o and radiology reports. A subset of reports was manually labeled to validate the GPT-4o labels. We then applied the pipeline to the resulting 9,459 brain MRIs from 1,137 patients at UCSF.

Results/Outcome

Segmentation median Dice score across an independent test set of 93 exams was 0.91 (figure 1). GPT-4o achieved >90% accuracy compared to manual labeling (figure 2). Longitudinal tumor volume charts with confidence intervals, segmentation overlays, and matched report labels were automatically created, highlighting changes visible only through multi-timepoint visualization.

Conclusion

This fully automated pipeline enables robust, confident tracking of meningioma volume across routine MRIs. By combining segmentation, registration, and interpretability, it facilitates more precise assessments of tumor progression than current single-prior comparisons allow. We are currently working towards internal clinical deployment of this tool for prospective analysis.

Statement of Impact

Our system shifts the paradigm from subjective pairwise comparison of meningiomas, the current clinical standard of care, to comprehensive quantitative and volumetric longitudinal analysis. By making multi-timepoint volume trends accessible and interpretable, this tool can reveal growth patterns that are otherwise easily missed –

improving surveillance strategies, surgical planning, and radiologic reporting.

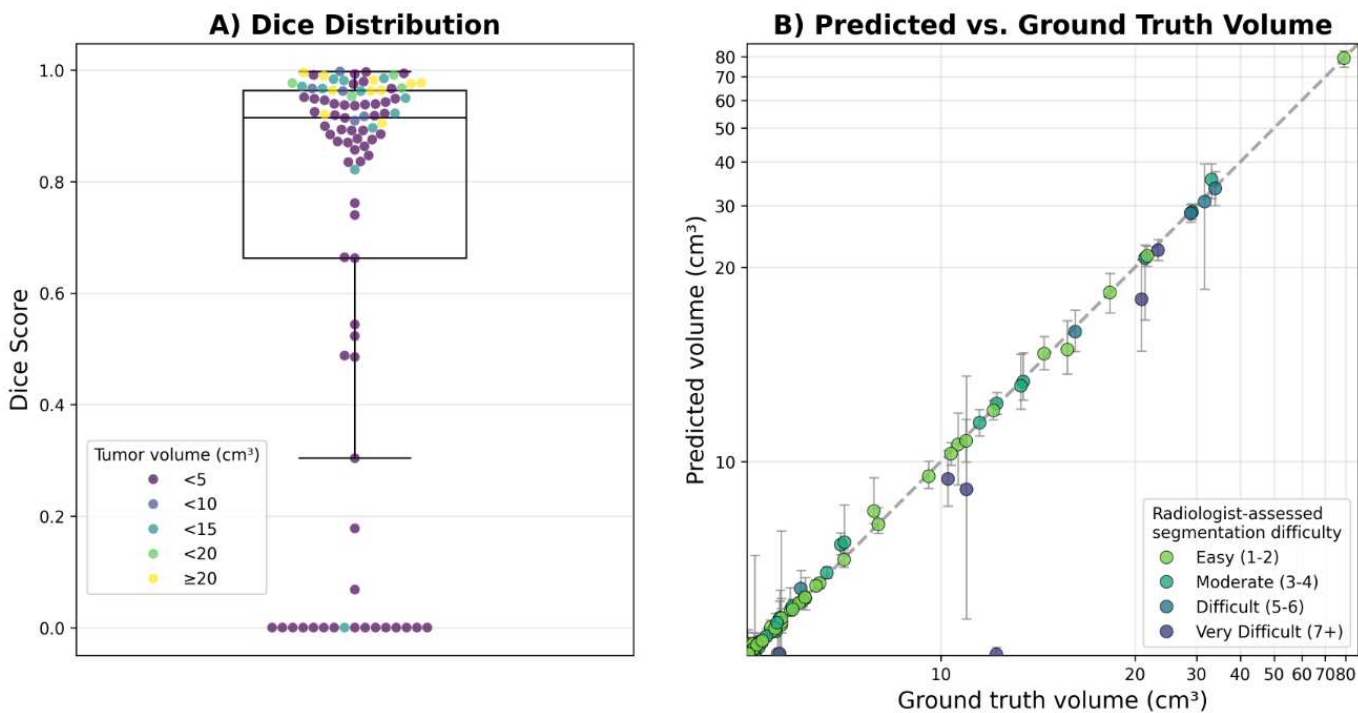


Figure 1: Model performance varies by tumor size and segmentation difficulty. A) Distribution of Dice scores across all cases, colored by ground truth tumor volume. Larger tumors (≥ 20 cm³, yellow) tend to achieve higher Dice scores, while smaller tumors (

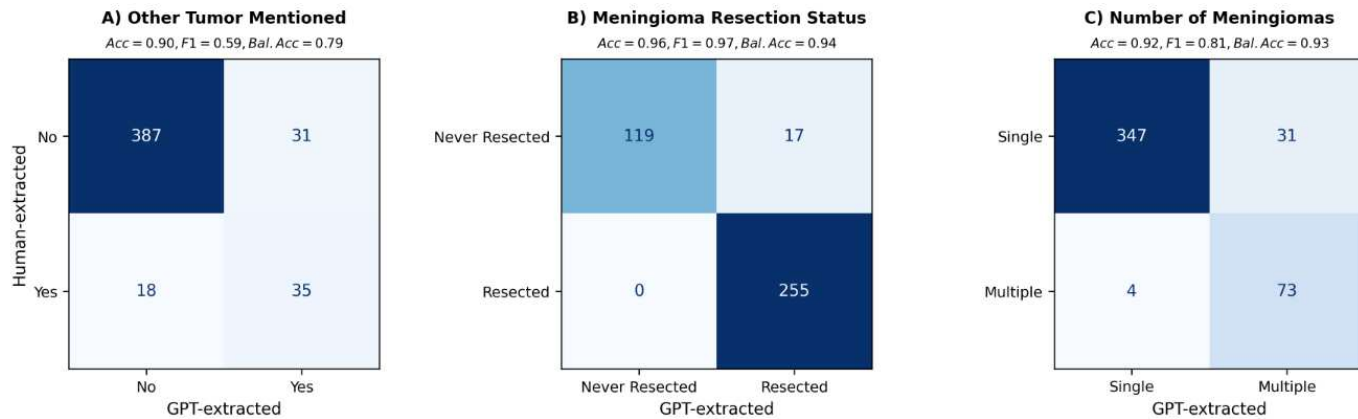
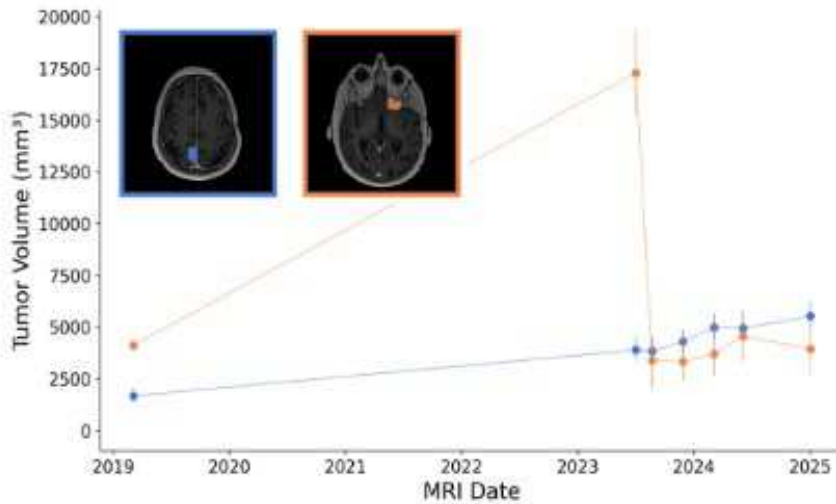


Figure 2: Multi-panel confusion matrices illustrating concordance between radiologist-annotated labels (y-axis) and PHI-compliant GPT-4o predictions (x-axis) for: (A) presence of other tumors, (B) meningioma resection status, and (C) meningioma count. Abbreviations: Acc = Accuracy, F1 = F1-score, Bal.Acc. = Balanced Accuracy

A) Meningioma segmentation volume over time



B) Segmentation overlays on T1 post contrast over time

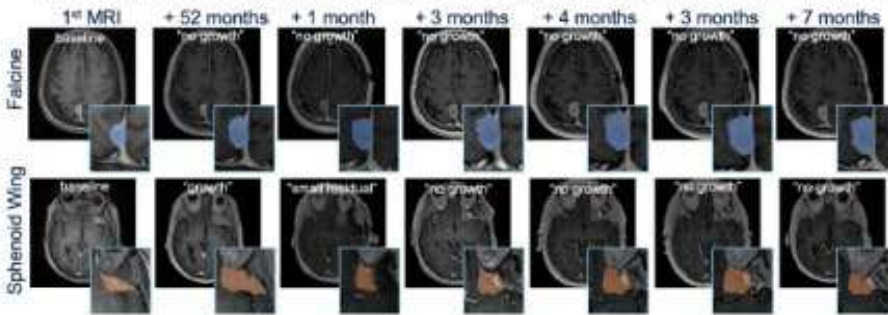


Figure 3: Example of meningioma tracking over time for a patient with (A) volume plot, confidence intervals, and 3D segmentations of two distinct meningiomas. This example demonstrates the slow growth of falcine meningioma (blue) and appropriately tracks residual sphenoid wing meningioma status post partial resection (orange). (B) Segmentation for each time point with radiologist assessment extracted from each exam's radiology report.

Keywords

Longitudinal Imaging; Volumetric Segmentation; Radiology AI; Confidence Scoring; Meningioma; Deep Learning