2026 JULY 19–22 VANCOUVER, BC
JOINT AAPM | COMP MEETING

Clinical Validation of an AI-Driven Detection Assistant for Gamma Knife Radiosurgery: Improving Workflow Confidence and Target Identification

Alexander Choi, MS[1]; Yazan Bakdash[2]; Poonam Yadav, PhD[1,2]
[1]Northwestern Memorial Hospital, [2]Northwestern University

ABSTRACT

Purpose: To evaluate the clinical feasibility of an in-house AI-driven detection tool for brain metastases in Gamma Knife radiosurgery. Small or interval metastases are prone to oversight during initial radiologic review yet require identification during the pre-treatment workflow. This study aims to deploy an in-house deep learning model specifically optimized to flag these subtle, small-volume lesions, acting as a safety filter to ensure comprehensive treatment delivery.

Methods: T1-post MRI sequence data from 368 patients were utilized to train a detection model using the nnU-Net framework, with a separate 93-patient dataset reserved for validation. To address the clinical challenge of identifying small lesions (<27 mm³), we compared a baseline model trained with standard Dice loss against a model utilizing TopK loss, designed to penalize hard-to-detect structures. Performance was evaluated primarily by detection sensitivity stratified by tumor size (small, medium, and large), reflecting the clinical priority of target identification over perfect geometric overlap.

Results: The baseline Dice loss model achieved a mean DSC of 0.82. Stratified sensitivity analysis showed robust detection for large (5 mm³, sensitivity 0.96) and medium tumors (0.88), but limited performance for small tumors (0.57). Implementation of the TopK loss function marginally improved overall DSC to 0.83 but yielded a marked improvement in small lesion detection, increasing sensitivity to 0.63 (10% improvement). Sensitivity for large (0.96) and medium (0.90) tumors remained high.

Conclusion: This study demonstrates the potential of an optimized in-house AI model to serve as an automated screening tool for Gamma Knife planning. While global DSC remained stable, the integration of TopK loss significantly enhanced sensitivity to small, easily overlooked tumors. By acting as a supplementary "second read," this tool streamlines metastasis identification, minimizing the risk of missed targets and ensuring the high precision required for effective patient outcomes.

CONTACT

Alexander Choi
Northwestern Memorial Hospital
251 E Huron St
Alexander.choi1@nm.org
615-584-4843

INTRODUCTION

Brain metastases are among the most common intracranial malignancies, with Gamma Knife radiosurgery (GKS) representing a cornerstone treatment modality. Comprehensive lesion identification prior to treatment is critical – each undetected metastasis represents a missed treatment target with direct implications for disease progression and patient outcomes.

Small lesions (< 27 mm^3) are disproportionately vulnerable to oversight during upstream radiologic review. By the time a patient reaches radiation oncology for planning, small or interval metastases may have been missed – with no systematic safety net to catch them.

Objective

No automated tool exists within the RO planning step specifically designed to flag small, easily overlooked lesions before treatment delivery begins.

Objective is to develop and validate an in-house deep learning segmentation model, optimized for sensitivity to small brain metastases, to serve as an automated second-read safety filter integrated into the pre-GKS planning workflow.

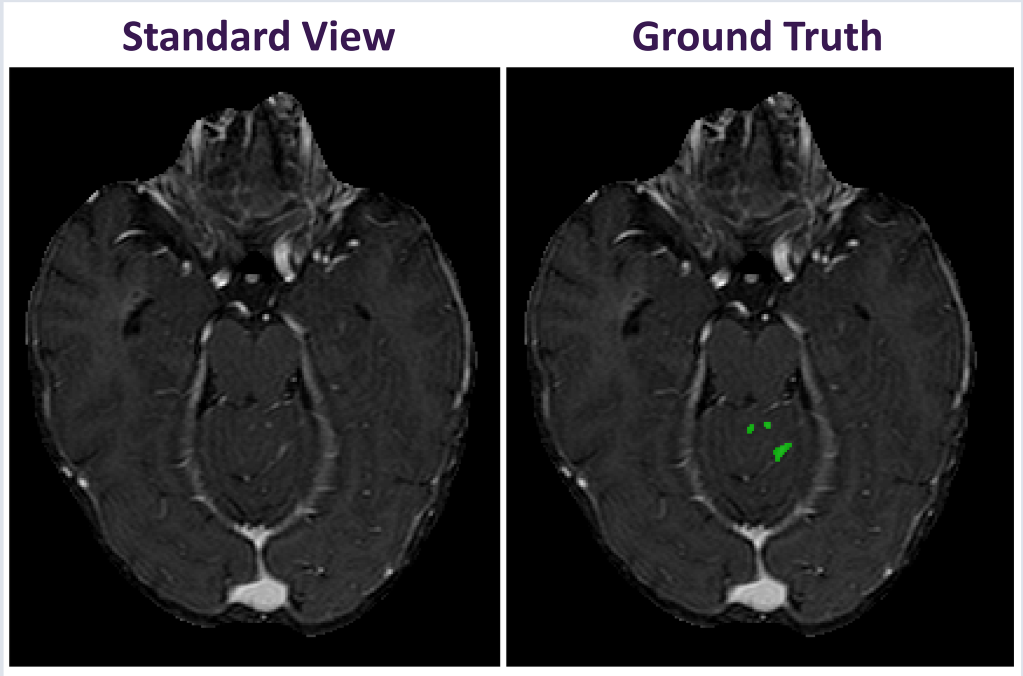

Figure 1. T1 post-contrast axial slice example with multiple lesions with volume <27 mm^3 ($7.1, 5.7, 22.7\ mm^3$)

METHODS AND MATERIALS

Dataset
- 368 patients for training, 93 holdout for validation
- Sequence: T1 post-contrast
- Ground-truth: Radiation oncologist-contoured GK treatment contour
- Single institution, retrospective

Multiple approaches were systematically evaluated before converging on an optimized solution (Table 2). nnU-Net framework provided standardized patch-based sampling, advanced augmentation suite, and automated inference, establishing a strong, reproducible baseline (Dice: 0.82). To improve sensitivity to overlooked small lesions, nnU-Net's loss was optimized by pairing the standard Dice loss with a Top-K Cross-Entropy loss. This compound formulation restricts voxel-level gradient computations strictly to the hardest 10% of voxels per patch during each training iteration, forcing the network to focus on the most challenging structures.

RESULTS

Detection sensitivity was evaluated stratified by lesion volume, reflecting the clinical priority of target identification over geometric overlap alone.

Key findings
- Both optimized loss functions achieved a statistically significant 10% relative improvement in small tumor detection sensitivity
- Global dice score was preserved (0.82 to 0.83), specificity not sacrificed
- Medium and large tumor sensitivity remained high (> 0.90, >0.96), and showed non-significant trends towards improvement
- Notably, instance-size weighting also improved edge delineation of medium and large tumors
- Secondary benefit of penalizing small boundary features

Table 1. Global Dice and Size-Stratified Detection Sensitivity

Model	Small	Medium	Large	Global Dice
nnU-Net Baseline	0.57	0.89	0.96	0.82
+Top-K Loss	0.63	0.90	0.96	0.83
+Instance-Size Weighting	0.63	0.91	0.96	0.83

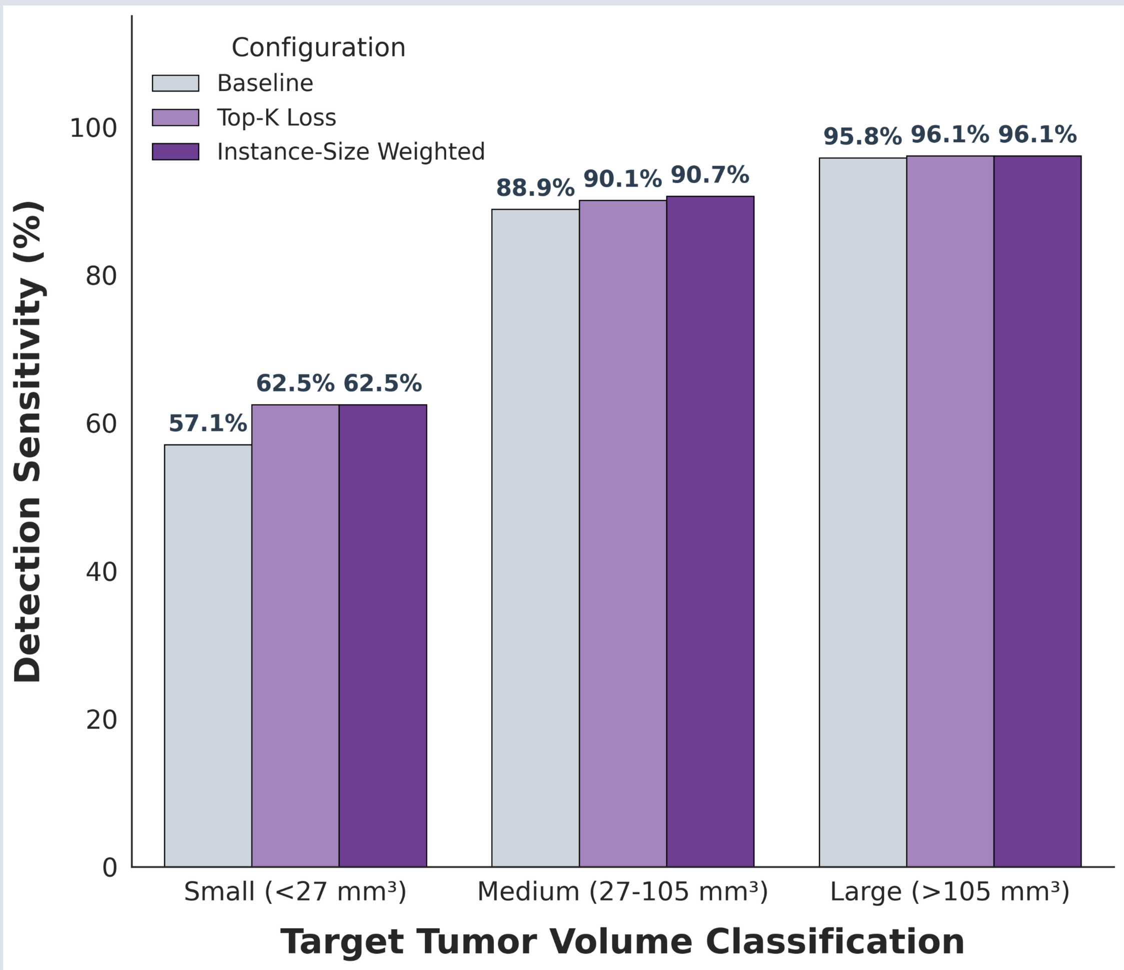

Figure 2. Size-stratified brain metastasis detection sensitivity across configurations

Table 2. Results and limiting factors of approaches during development

Approach	Result	Limiting Factor
3D U-Net	Dice: 0.42	Lacked patch sampling, augmentation
DSIT U-Net	Failed to replicate literature	Architectural gains not reproducible
SwinUNetR	Dice: 0.82	Lacks CNN spatial bias
3D medYOLO	Poor precision/recall	Aggressive downsampling
Class-split CE	Performance decline	Boundary confusion penalty

CONCLUSIONS

This work demonstrates that targeted loss function optimization rather than architectural complexity is the key level for improving small brain metastasis detection in a limited single-institution dataset. Both Top-K voxel clipping and voxel-wise instance-size weighting achieved statistically significant improvements in small lesion sensitivity, the primary clinical success criterion for GK pre-treatment safety filter.

By acting as an automated second-read screening step before treatment planning, this tool has the potential to reduce missed targets and improve the comprehensiveness of GK radiosurgery delivery, directly supporting patient outcomes.

Future direction
Dataset expansion, prospective clinical evaluation, and integration into the pre-GK planning workflow as real-time detection assistant.

Acknowledgements
All data collection and usage were IRB-approved

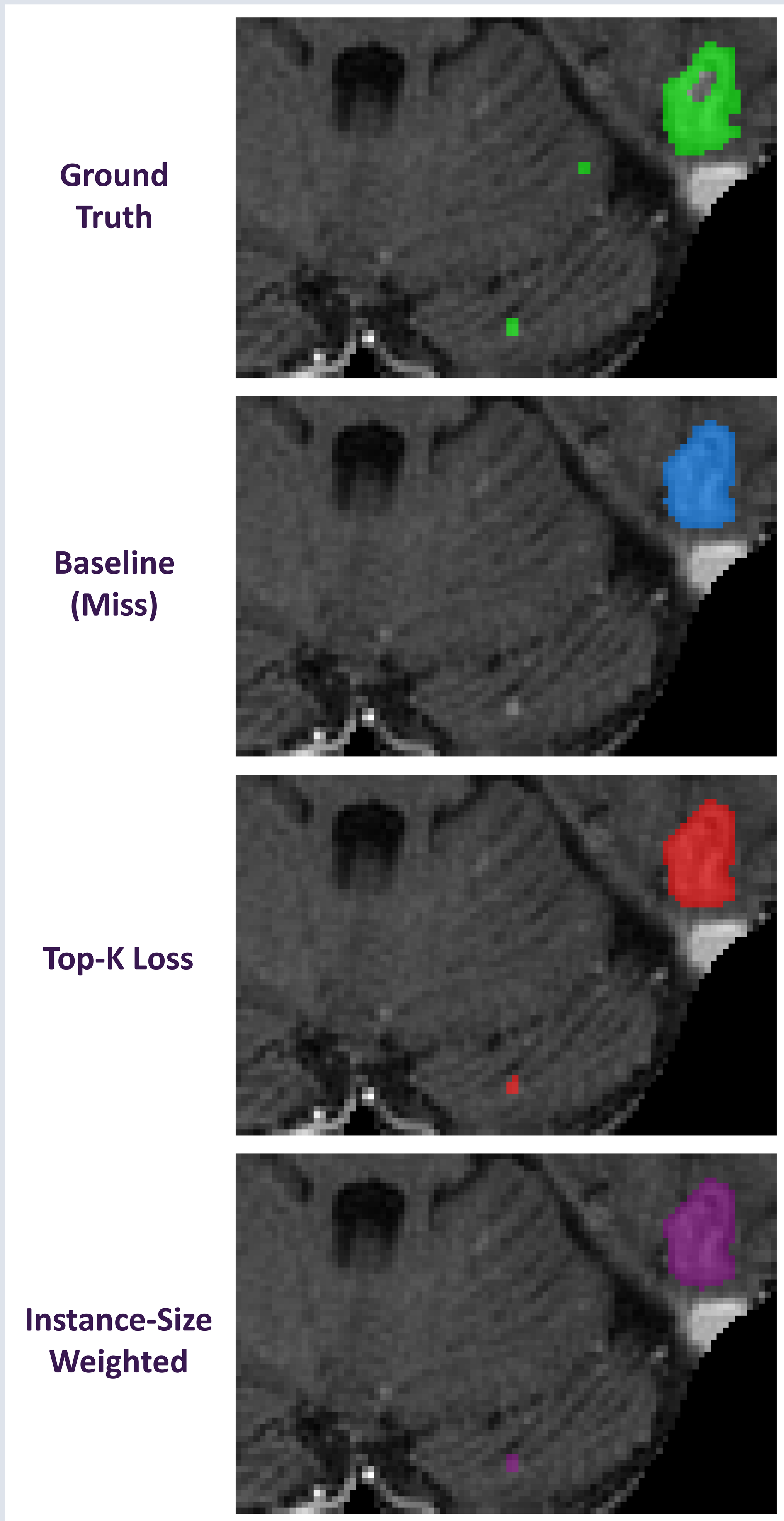

Figure 3. Detection layout of an overlooked small metastasis (7.91 mm³)

Northwestern Medicine®