This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Integrating nnDetection and nnU-Net for Brain Metastasis Segmentation in the BraTS-METS 2026 Challenge

Jun-Ho Kim[1(✉)*] [0000-0003-0432-9009] and Ki Bok Nam[1*][0000-0003-0317-1860]

[1] Neuro AI Division DEEPNOID Inc., Seoul, Republic of Korea
rla803wns@gmail.com

Abstract. Brain metastases are the most frequently diagnosed intracranial tumors in adults, and their accurate detection and volumetric segmentation are essential for treatment planning and response assessment. This task is challenging because metastatic lesions are often small and numerous, making them easy to miss while also being prone to false-positive predictions. We address the BraTS 2026 Brain Metastases challenge with a pipeline that couples a segmentation branch and a detection branch. The segmentation branch, built on nnU-Net, comprises a region-based primary model and a dedicated resection-cavity (RC) specialist. Their predictions are combined by a priority-based merging rule to produce the multi-class segmentation, and a size-based post-processing then removes connected components smaller than 5 mm^3. In parallel, the detection branch, an nnDetection model trained on the enhancing tumor (ET), provides an independent second opinion. A detection-guided consensus keeps a small enhancing-tumor component (smaller than 27 mm^3) only when its centroid falls within a detection box, suppressing spurious small detections while preserving lesions corroborated by the detector. On the challenge validation set, our full pipeline (configuration 5) achieved lesion-wise Dice similarity coefficients (DSC) of 0.745, 0.671, 0.765, and 0.720 for the ET, RC, tumor core (TC), and whole tumor (WT), respectively, with corresponding normalized surface distances (NSD) of 0.808, 0.547, 0.816, and 0.735. Relative to the segmentation-only configurations, using the detector to prune false-positive small detections yielded the highest values for all these overlap metrics, at small cost to the small-lesion detection F1-score (e.g., 0.501 for ET). Our code is available at https://github.com/deepnoid-ai/DEEP-MET.

Keywords: Brain metastases · Detection · Segmentation · Deep learning

1 Introduction

Brain metastases are the most frequently diagnosed intracranial tumors in adults, arising in an estimated 20–40% of patients with systemic cancer, most often from lung, breast, and colon primaries [1,2]. Longer survival under modern systemic therapy and more

* Jun-Ho Kim and Ki Bok Nam contributed equally to this work

sensitive imaging have raised their reported incidence, yet prognosis remains guarded, placing a growing burden on neuro-oncology workflows [2,3]. Monitoring is labor-intensive, especially with multiple lesions: although the Response Assessment in Neuro-Oncology Brain Metastases (RANO-BM) guidelines rely on the largest unidimensional diameter, volumetric assessment of the tumor and surrounding edema is increasingly recognized as critical for treatment planning, particularly stereotactic radiosurgery, and for reliable response assessment [4,5]. Performed manually, these tasks are subjective, rater-dependent, and not scalable, which motivates automated, reproducible solutions [6].

Automated detection and segmentation are complicated by the small size and multiplicity of the lesions. Early template-matching methods [7] gave way after 2018 to deep learning, driven by larger and more heterogeneous datasets [8]; encoder–decoder architectures became standard [9], and the self-configuring nnU-Net established the state of the art across a broad range of segmentation tasks [10]. To benchmark such methods on standardized, multi-institutional data, the Brain Tumor Segmentation (BraTS) challenge introduced a dedicated brain-metastasis track, BraTS-METS [11], whose 2025 Lighthouse edition extended it to pre- and post-treatment and longitudinal multiparametric MRI (mpMRI) acquired with diverse scanners and protocols [12]. This heterogeneity makes BraTS-METS a demanding testbed for methods intended for clinical deployment. Within BraTS-METS specifically, nnU-Net has formed the backbone of the strongest submissions, and the winning solution of the BraTS-METS 2025 challenge was built entirely on nnU-Net, varying the loss function, model size, training duration, and post-processing, and reaching a Dice score above 0.78 for the enhancing tumor (ET) with a Residual Encoder (ResEnc) configuration [13].

A recurring difficulty is the handling of small metastatic lesions, which are easily missed yet also prone to false-positive predictions. This is especially consequential under the challenge's lesion-wise evaluation, where every unmatched predicted component counts as a zero-overlap lesion, so even small false positives penalize the overlap metrics and not only the detection score. We therefore propose, for the BraTS 2026 Metastases challenge, a pipeline that couples segmentation with dedicated detection: an nnU-Net produces the multi-class segmentation, a lightweight post-processing discards connected components smaller than 5 mm^3 per class, and an nnDetection model then removes the remaining small false positives while retaining as many true lesions as possible. On the challenge dataset, suppressing small false positives in this way improves the lesion-wise overlap metrics, the Dice similarity coefficient (DSC) and the normalized surface distance (NSD), at only a modest, controllable cost to small-lesion detection.

This article is structured as follows. Section 2 presents the proposed method, covering the dataset, the framework, the network configuration and training, and the evaluation metric used within the challenge. The experimental results are reported in Section 3 and discussed in Section 4. Our code is publicly available at: https://github.com/deepnoid-ai/DEEP-MET.

2 Methods

2.1 Dataset

We use the dataset released through the BraTS 2026 Brain Metastases challenge [12], a retrospective, multi-institutional collection of pre- and post-treatment brain metastasis mpMRI scans acquired under standard clinical conditions with diverse scanners and protocols. The 2026 edition builds on the 2025 dataset, with all released cases annotated and an additional quality-control pass. Each study provides four co-registered sequences, namely pre-contrast T1-weighted (T1W), post-contrast T1-weighted (T1C), T2-weighted (T2W), and T2 fluid-attenuated inversion recovery (FLAIR), and ground-truth masks follow a four-label protocol: non-enhancing tumor core (NETC), surrounding non-enhancing FLAIR hyperintensity (SNFH), ET, and resection cavity (RC), the last for post-treatment cases. Masks were produced by Simultaneous Truth and Performance Level Estimation (STAPLE) fusion of several algorithms, manually refined by neuroradiology experts, and approved by board-certified neuroradiologists. Contributed by eight institutions, the dataset is split into 1,296 training, 179 validation, and 303 testing cases; 1,272 (from the University of California, San Francisco and San Diego) are provided in native space co-registered to T1C at 1 mm³ and 506 in the SRI24 atlas space, the native cases avoiding interpolation that can distort images and obscure small lesions.

2.2 Framework

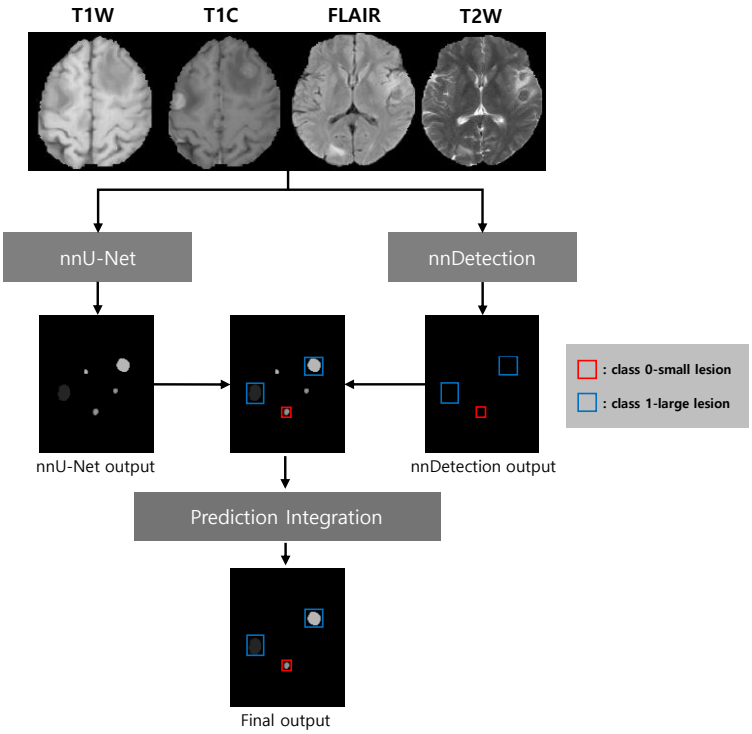

Fig. 1. Overview of the framework. All four MRI modalities (T1W, T1C, FLAIR, T2W) are fed to both an nnU-Net branch (multi-class segmentation) and an nnDetection branch, which outputs bounding boxes around detected ET (red: small, class 0; blue: large, class 1). In the prediction integration, predicted large components are always kept, while small components are retained only if their centroid lies within a detection box and discarded otherwise.

Our framework combines two complementary self-configuring models, a segmentation branch and a detection branch, whose predictions are merged in a prediction integration step (Figure 1). Both branches receive the same four co-registered mpMRI modalities (T1W, T1C, T2W, and FLAIR) as input.

Segmentation branch (nnU-Net). This branch performs voxel-wise multi-class segmentation of NETC, SNFH, ET, and RC using nnU-Net, combining two models by a priority-based merging rule: a region-based primary model and a dedicated RC specialist. The primary model uses nnU-Net's region-based formulation, in which the four labels are converted into nested regions defined to match the challenge's evaluation targets: ET, tumor core (TC), whole tumor (WT), and RC. Because RC was the weakest sub-region, particularly for small, isolated instances, we additionally trained a binary (background-vs.-RC) specialist restricted to RC-positive cases, otherwise identical to the primary model. At inference, the two are merged at the connected-component level: a specialist RC component is inserted only where it lies entirely within primary-model background, so the specialist can recover missed RC lesions without overwriting existing predictions. A size-based post-processing is then applied to the merged output, removing connected components smaller than 5 mm^3 in every class. This step is first introduced in configuration 3 and retained in all subsequent configurations (Table 1).

Detection branch (nnDetection). In parallel, we use nnDetection [14], a self-configuring framework for medical object detection, to localize lesions at the object level and output bounding boxes. Because metastatic lesions are predominantly characterized by their ET, we train nnDetection to detect the ET component only. Each detection is assigned to one of two size classes defined by the ET volume: small (<27 mm^3; class 0) and large (≥ 27 mm^3; class 1). This 27 mm^3 boundary coincides with the challenge's small-lesion evaluation threshold, so the detector is explicitly tuned to the regime in which detection, rather than boundary overlap, governs the score. Because the dataset is provided in two registration spaces, we train two separate nnDetection models, one on the native-space cases and one on the SRI24-space cases, and apply each model to inputs in its corresponding space. The two spaces can be identified automatically, since SRI24-registered volumes always have a fixed size of $240 \times 240 \times 155$ voxels, whereas native-space volumes do not.

Prediction integration. The two branches are integrated to form the final prediction (Figure 1). We first apply the size-based post-processing introduced above to the merged segmentation output, after which the detection-guided consensus is applied to the ET only, since nnDetection is trained to detect ET: each ET component with a volume of at least 27 mm^3 is kept, whereas each ET component smaller than 27 mm^3 is kept only if its centroid lies within a small-lesion nnDetection box and is discarded otherwise. Components of the other classes (NETC, SNFH, RC) are not modified by this step. This selectively prunes small ET false positives while preserving the small ET lesions corroborated by the detector.

2.3 Network Configuration and Training

All models in both branches were trained on a NVIDIA Quadro RTX 8000 GPU. Both models use nnU-Net's default preprocessing and training schedule (Table 1); we depart from the defaults only in the augmentation and the loss. Following nnU-Net's automatic configuration, the primary and RC-specialist models used 3D full-resolution patch sizes of 160 × 224 × 192 and 224 × 256 × 224 voxels, with batch sizes of 4 and 2, respectively. Besides nnU-Net's default spatial, Gaussian noise, and blur transforms, we apply a mask-based local intensity augmentation adapted from the local location-scale augmentation (LLA) of Saliency-balancing Location-scale Augmentation (SLAug) [15]: the label map partitions each patch into regions, each undergoing an independently sampled affine transform $x' = a \cdot x + b$, improving robustness to appearance shifts across scanners and sites. The loss combines a cross-entropy term with a recall-oriented Tversky term [16]. For the primary model this cross-entropy term is a binary cross-entropy (BCE) over the region channels, whereas for the specialist it is a standard cross-entropy. The Tversky term uses an asymmetric weighting that favors recall ($\delta = 0.7$), and deep supervision is applied at all decoder resolutions. Both models are trained for 2,000 epochs on the "all" fold.

The nnDetection models were trained using their default, self-configured settings, without any manual modification to the automatically determined preprocessing, network topology, or training schedule [14]. Following this automatic configuration, the native-space and SRI24-space models used patch sizes of 80 × 160 × 128 and 64 × 64 × 56 voxels, respectively, both with a batch size of 4.

The 5 mm^3 component-removal threshold was not tuned by any automated procedure but set heuristically. We evaluated several candidate thresholds on the challenge validation set and selected 5 mm^3 as the value that best suppressed spurious small components while preserving genuine lesions under the lesion-wise metric.

The detection score thresholds were set in the same way. We evaluated several candidate values on the challenge validation set and selected 0.15 for the native space and 0.2 for the SRI24 space as the values that gave the best balance between the overlap metrics and the small-lesion detection score. The two registration spaces were tuned independently, since they are handled by separate detectors whose confidence distributions need not coincide.

Table 1. Summary of the five configurations evaluated in this study, forming an incremental ablation in which each step adds a single component; each configuration is identified by its submission ID. Configurations 1 to 4 use nnU-Net alone, whereas configuration 5 additionally integrates nnDetection through the prediction integration. All configurations share the same settings, namely the nnUNetPlannerResEncXL planner (3d_fullres), the "all" training fold (nnU-Net's setting that uses all training cases without a held-out fold), an initial learning rate of 0.01, Gaussian noise-and-blur augmentation, and test-time augmentation (TTA) with mirroring and Gaussian transforms; the table lists only the components that differ across configurations. For the nnDetection integration in configuration 5, the detection score thresholds were 0.15 (native space) and 0.2 (SRI24 space). ✓ indicates that a component is enabled and – that it is not used.

ID (brats id)	LLA	Loss	Post-proc	RC specialist	nnDetection

			(<5mm³)		
1 (9767797)	–	BCE+DSC	–	–	–
2 (9768440)	✓	BCE+Tversky	–	–	–
3 (9770484)	✓	BCE+Tversky	✓	–	–
4 (9771732)	✓	BCE+Tversky	✓	✓	–
5 (9772816)	✓	BCE+Tversky	✓	✓	✓

2.4 Evaluation Metrics

To quantify the quality of predictions, the following metrics are computed lesion-wise for each region of interest, after decomposing predictions and references into connected components and matching them by Dice overlap. Reference lesions are split into large and small by a volume threshold of 27 mm³:

- **Dice Similarity Coefficient (DSC)**, the overlap metric calculated between the predicted and ground-truth region of each large lesion (≥ 27 mm³), ranging from zero to one, with one indicating the perfect score, i.e., full agreement with the ground truth.
- **Normalized Surface Distance (NSD)**, the boundary agreement metric calculated between the predicted and ground-truth surfaces of each large lesion, with one indicating the perfect score.
- **F1-score**, calculated for small lesions (< 27 mm³) as $2 \cdot TP / (2 \cdot TP + FP + FN)$, where TP, FP, and FN denote true positives, false positives, and false negatives, respectively, and a matched lesion is counted as a true positive when its per-lesion DSC exceeds 0.1.

While calculating the quality metrics, the following regions of interest are considered, and each metric is calculated for each region separately: (i) ET, (ii) TC, (iii) WT, and (iv) RC. The resulting region-wise metrics are then aggregated across all cases to rank the participating methods.

3 Results

Table 2. Validation performance of the five submitted configurations, each identified by its submission ID, reported per evaluation region. For every configuration we list the DSC and the NSD for the ET, RC, TC, and WT. For each metric, the best value across configurations is shown in bold.

ID (brats id)	ET_{dsc}	ET_{nsd}	RC_{dsc}	RC_{nsd}	TC_{dsc}	TC_{nsd}	WT_{dsc}	WT_{nsd}
1 (9767797)	0.665	0.728	0.575	0.494	0.693	0.742	0.662	0.683
2 (9768440)	0.682	0.745	0.543	0.415	0.711	0.762	0.695	0.714
3 (9770484)	0.731	0.793	**0.671**	**0.547**	0.751	0.803	0.718	0.734
4 (9771732)	0.731	0.793	**0.671**	**0.547**	0.751	0.803	0.718	0.734
5 (9772816)	**0.745**	**0.808**	**0.671**	**0.547**	**0.765**	**0.816**	**0.720**	**0.735**

Table 3. Small-lesion detection performance of the five configurations, reported as the F1-score for lesions smaller than 27 mm³, given separately for the ET, RC, TC, and WT regions. For each region, the best F1-score across configurations is shown in bold.

ID (brats id)	ET_{F1}	RC_{F1}	TC_{F1}	WT_{F1}
1 (9767797)	0.476	0.000	0.486	0.388
2 (9768440)	0.492	0.111	0.506	0.406
3 (9770484)	**0.511**	0.000	**0.518**	**0.411**
4 (9771732)	**0.511**	**0.167**	**0.518**	**0.411**
5 (9772816)	0.501	**0.167**	0.507	0.402

Table 2 reports the lesion-wise DSC and NSD for each evaluation region, and Table 3 reports the F1-score for lesions smaller than 27 mm³.

On the overlap metrics (Table 2), configuration 5 obtained the best scores for the ET (DSC 0.745, NSD 0.808), the TC (DSC 0.765, NSD 0.816), and the WT (DSC 0.720, NSD 0.735), and it matched the best RC score (DSC 0.671, NSD 0.547), which it shared with configurations 3 and 4. The nnU-Net-only configurations were consistently lower on these regions: configurations 1 and 2 trailed the most (ET DSC 0.665 and 0.682, respectively), while configurations 3 and 4 were intermediate and identical to each other (ET DSC 0.731, TC DSC 0.751, WT DSC 0.718).

On small-lesion detection (Table 3), the ranking differed. Configurations 3 and 4 achieved the highest F1 for the ET (0.511), the TC (0.518), and the WT (0.411), with configuration 5 marginally lower on all three (0.501, 0.507, and 0.402, respectively). For the RC, configurations 4 and 5 shared the best F1 (0.167), whereas configurations 1 and 3 scored 0. Overall, integrating nnDetection (configuration 5) produced the strongest overlap metrics while giving up a small amount of small-lesion F1 relative to the best nnU-Net-only configurations.

The effect of each component is also visible qualitatively (Figure 2). In Figure 2a, the predicted segmentations change progressively from configuration 1 to configuration 3 in all regions except the RC, whereas the RC is more affected by the later components. In Figure 2b, the RC specialist (configuration 4) adds small RC components that are absent from configuration 3. As no ground truth is available for these validation cases, these panels illustrate the difference between configurations rather than confirm their correctness, which is instead quantified in Tables 2 and 3.

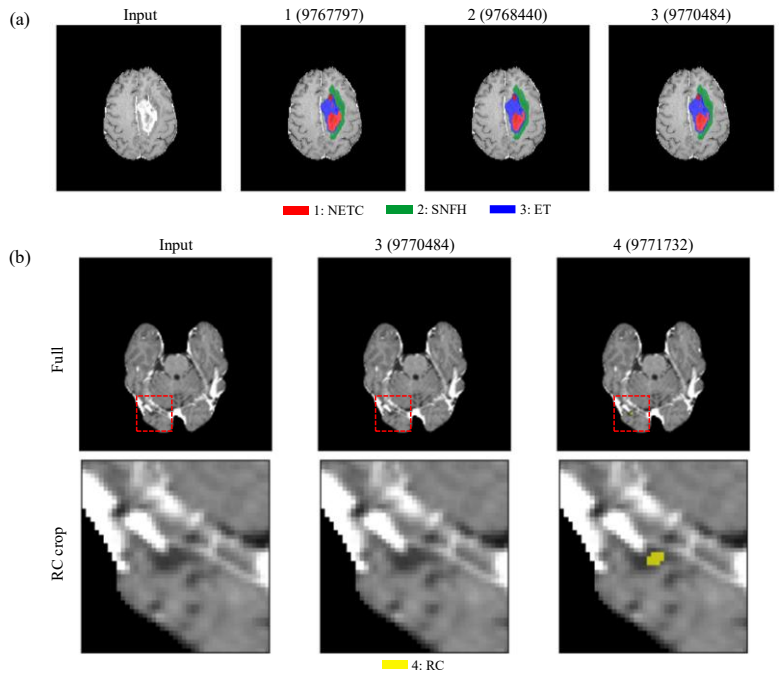

Fig. 2. Qualitative comparison of predicted segmentations on validation cases (no ground truth available), overlaid on the T1C image. (a) Configurations 1, 2, and 3, showing how the predictions change across all regions except the RC. (b) Configurations 3 and 4, where the RC specialist adds small RC components not present in configuration 3.

Figure 3 shows a representative validation case, for which no ground truth is available. The two configurations produce identical segmentations except for a few small ET components that the detection-guided consensus removed in configuration 5 (red circles). Because these cases are unlabeled, we cannot determine on a per-component basis whether each removal was a false positive or a true lesion. Their aggregate effect, however, is measurable: across the validation set, the consensus substantially increased the lesion-wise DSC and NSD while only slightly decreasing the small-lesion F1 (Tables 2 and 3).

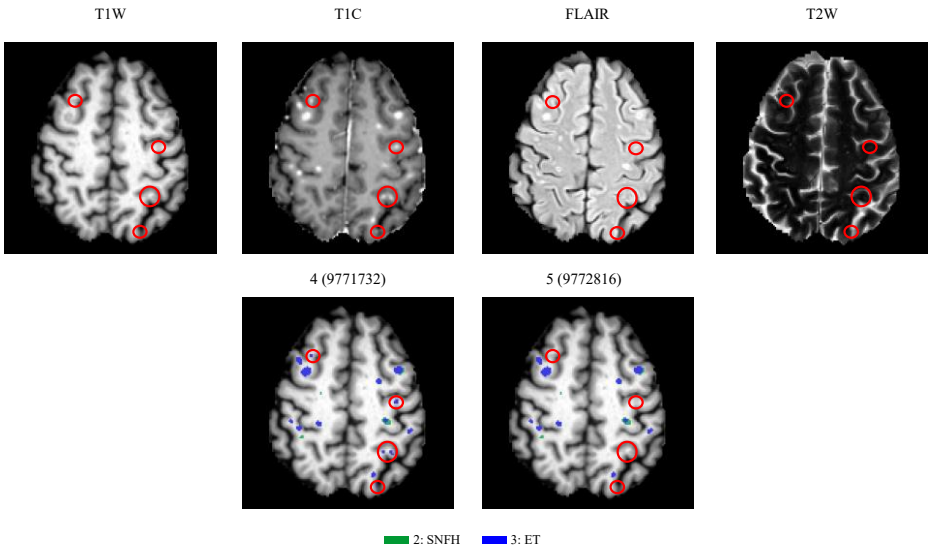

Fig. 3. Qualitative example on a validation case (no ground truth available). Top row: the four input mpMRI modalities (T1W, T1C, FLAIR, T2W). Bottom row: predictions of configuration 4 (nnU-Net only) and configuration 5 (with nnDetection), overlaid on the T1C image, with ET in blue and SNFH in green. Red circles mark the small ET components removed by the detection-guided consensus in configuration 5; whether each was a false positive or a true lesion cannot be verified without ground truth.

Because no ground truth is available for the challenge validation set, the preceding figures can only show how the predictions differ between configurations, not which prediction is correct. To verify the effect of the detection-guided consensus against reference annotations, we additionally partitioned the annotated training set into a training subset and a held-out labeled test subset and repeated the analysis there. Figure 4 shows an example from this labeled subset in which a small ET component predicted by configuration 4 is absent from the reference annotation and is removed by the consensus in configuration 5, confirming that the removed component was indeed a false positive.

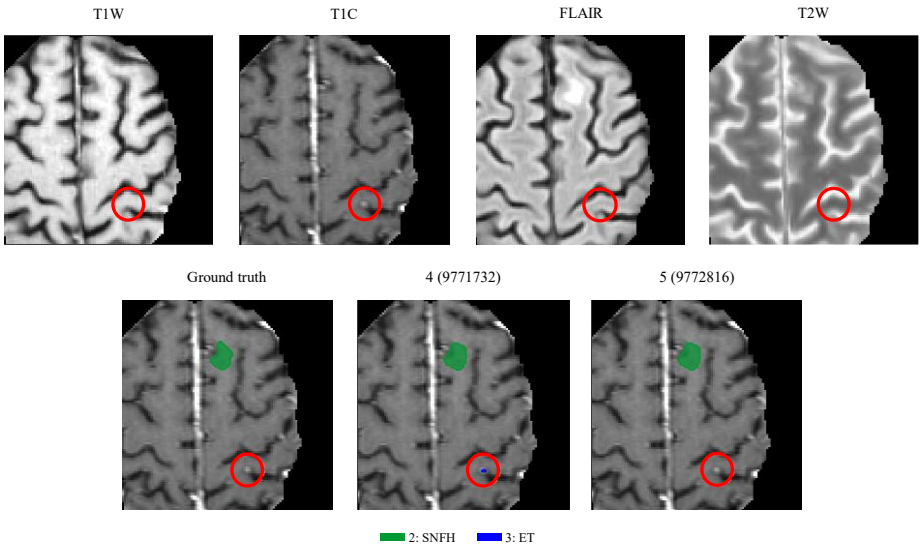

Fig. 4. Qualitative validation of the detection-guided consensus on a held-out labeled subset of the training set, for which ground truth is available. Top row: the four input mpMRI modalities (T1W, T1C, FLAIR, T2W). Bottom row: the ground truth, the prediction before the consensus (configuration 4), and the prediction after it (configuration 5), overlaid on the T1C image, with ET in blue and SNFH in green. The red circle indicates the same location in all panels and marks a small ET component that is predicted by configuration 4, is absent from the ground truth, and is correctly removed by the consensus in configuration 5.

4 Discussion

For the segmentation branch, we build the final configuration incrementally, adding one component at a time (Tables 2 and 3). The mask-based LLA and Tversky loss (configuration 2) target domain shift and bias the objective toward higher recall, since the lesion-wise evaluation penalizes missed lesions more heavily than comparable false positives. They improved the ET, TC, and WT metrics, but lowered the RC overlap, suggesting that the more recall-oriented training introduced additional spurious RC predictions.

Integrating the detection branch (configuration 5) completed the pipeline, giving the best overlap metrics at a small cost to detection F1. This follows from lesion-wise scoring: every unmatched predicted component counts as a zero-overlap lesion, so the small ET false positives that nnU-Net occasionally produces depress DSC and NSD, not only the detection score. Acting as an independent second opinion, the detector removes the unconfirmed small ET components, precisely the most penalizing predictions, so configuration 5 improves the overlap metrics over configuration 4 without changing the underlying segmentation. We kept the consensus subtractive: the detector only removes components and never adds lesions. A union would raise small-lesion recall but reintroduce false positives, whereas the subtractive rule trades a little

detection F1 for a larger overlap gain, which is favorable under an aggregate ranking. The residual F1 cost reflects detector false negatives, since unconfirmed true small lesions are discarded, so the detector's recall bounds how many survive.

Our design has limitations that motivate future work. The RC specialist recovers only RC lesions missed entirely by the primary model and cannot refine partially overlapping ones, and because it is trained only on RC-positive cases it does not observe the RC-negative distribution and may add false positives, which we mitigate only indirectly through the background-restricted merging and the size-based post-processing. On the detection side, nnDetection was used with its default settings, leaving its potential unexplored, and the fusion is a fixed, purely subtractive centroid-in-box rule that is sensitive to the confidence threshold. Future work will refine the detector (its configuration, loss, and confidence calibration), add RC-negative cases to specialist training, and explore a more integrated, learned or probabilistic fusion, including an additive mode that can also propose lesions.

Acknowledgments. This work was supported by the Technology Innovation Program (RS-2025-02221011, Development of Medical-Specialized Multimodal Hyperscale Generative AI Technology for Global Integration) funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea).

Disclosure of Interests. Authors are employed by DEEPNOID Inc., a company that develops commercial medical AI software for medical image analysis.

References

1. Lamba, N., Wen, P.Y., Aizer, A.A.: Epidemiology of brain metastases and leptomeningeal disease. Neuro-Oncology 23(9), 1447–1456 (2021).
2. Ostrom, Q.T., Wright, C.H., Barnholtz-Sloan, J.S.: Brain metastases: epidemiology. In: Handbook of Clinical Neurology, vol. 149, pp. 27–42. Elsevier (2018).
3. Yri, O.E., et al.: Survival and quality of life after first-time diagnosis of brain metastases: a multicenter, prospective, observational study. Lancet Reg. Health Eur. 49, 101181 (2025).
4. Mansouri, A., et al.: Stereotactic radiosurgery for patients with brain metastases: current principles, expanding indications and opportunities for multidisciplinary care. Nat. Rev. Clin. Oncol. 22(5), 327–347 (2025).
5. Aizer, A.A., et al.: Brain metastases: a Society for Neuro-Oncology (SNO) consensus review on current management and future directions. Neuro-Oncology 24(10), 1613–1646 (2022).
6. Zhou, Z., et al.: MetNet: computer-aided segmentation of brain metastases in post-contrast T1-weighted MRI. Radiother. Oncol. 153, 189–196 (2020).
7. Ambrosini, R.D., Wang, P., O'Dell, W.G.: Computer-aided detection of metastatic brain tumors using automated three-dimensional template matching. J. Magn. Reson. Imaging 31(1), 85–93 (2010).
8. Cho, S.J., et al.: Brain metastasis detection using machine learning: a systematic review and meta-analysis. Neuro-Oncology 23(2), 214–225 (2021).
9. Grøvik, E., et al.: Deep learning enables automatic detection and segmentation of brain metastases on multisequence MRI. J. Magn. Reson. Imaging 51, 175–182 (2020).
10. Isensee, F., et al.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat. Methods 18(2), 203–211 (2021).

11. Moawad, A.W., et al.: The Brain Tumor Segmentation (BraTS-METS) Challenge 2023: Brain Metastasis Segmentation on Pre-treatment MRI. arXiv:2306.00838 (2024).
12. Maleki, N., et al.: Analysis of the MICCAI Brain Tumor Segmentation – Metastases (BraTS-METS) 2025 Lighthouse Challenge: Brain Metastasis Segmentation on Pre- and Post-treatment MRI. arXiv:2504.12527 (2025).
13. Bancerek, M., Rudzki, P., Nalepa, J.: Segmentation of Pre and Post-treatment Brain Metastases Using nnU-Nets. In: Bakas, S., et al. (eds). MICCAI 2025. LNCS, vol. 16376, pp. 161-172. Springer, Cham (2026). https://doi.org/10.1007/978-3-032-16365-3_15
14. Baumgartner, M., Jäger, P.F., Isensee, F., Maier-Hein, K.H.: nnDetection: A Self-configuring Method for Medical Object Detection. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12905, pp. 530–539. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87240-3_51
15. Su, Z., et al.: Rethinking data augmentation for single-source domain generalization in medical image segmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 37(2), pp. 2366-2374 (2023). https://doi.org/10.1609/aaai.v37i2.25332
16. Salehi, S.S.M., Erdogmus, D., Gholipour, A.: Tversky loss function for image segmentation using 3D fully convolutional deep networks. In: Wang, Q., Shi, Y., Suk, H.I., Suzuki, K. (eds.) MLMI 2017. LNCS, vol. 10541, pp. 379-387. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67389-9_44