Brain Metastasis Segmentation Using an Ensemble of U-Mamba Models

Piotr Rudzki[1*], Maria Bancerek[1*][0000−0001−7061−7020], and Jakub Nalepa[2,1][0000−0002−4026−1569]

[1] Graylight Imaging, Gliwice, Poland
{prudzki, mbancerek, jnalepa}@graylight-imaging.com
[2] Silesian University of Technology, Gliwice, Poland
Jakub.Nalepa@polsl.pl

Abstract. Accurate detection and segmentation of brain metastases are challenging tasks that are critical for quantitative imaging assessment and clinical decision-making. Current clinical evaluation relies primarily on manual measurements following the Response Assessment in Neuro-Oncology Brain Metastases (RANO-BM) guidelines, making the process labor-intensive, time-consuming, and prone to inter-/intra-observer variability and human bias. In this work, we extend our previous method (the winning solution of BraTS-METS 2025) by developing an ensemble framework based on the U-Mamba models. Building upon the nnU-Net framework, U-Mamba integrates convolutional neural networks with state space sequence models, enabling simultaneous extraction of local image features and modeling of long-range spatial dependencies. We evaluated multiple U-Mamba variants and post-processing strategies, and the final ensemble combines two models, with one model specifically optimized for small lesion detection through the use of the CATMIL loss function. The results returned by the evaluation server for the validation set of BraTS-METS 2026 are (Dice Similarity Coefficient/Normalized Surface Distance): 0.708/0.767, 0.499/0.427, 0.740/0.789, and 0.694/0.711 for the enhancing tumor, resection cavity, tumor core, and whole tumor. The detection (F1 score) metrics are: 0.450, 0.457, and 0.357 for enhancing tumor, tumor core, and whole tumor.

Keywords: Brain metastasis · Detection · Segmentation · nnU-Net · U-Mamba · State space sequence model.

1 Introduction

Monitoring metastatic brain lesions is a costly and time-consuming task, especially in patients with multiple metastases, where assessments are based on manual evaluation. Therefore, this procedure can be subjective, prone to human errors and bias, and dependent on the experience of a particular reader. According to the Response Assessment in Neuro-Oncology Brain Metastases

* Joint first author with equal contribution.

(RANO-BM) guidelines, brain metastases are conventionally assessed using the largest unidimensional diameter of each lesion. Nevertheless, accurate volumetric quantification of both metastatic lesions and the associated peritumoral edema is essential for supporting clinical decision-making and improving predictions of treatment response. This challenge is further exacerbated by the typically small size of brain metastases, which makes their detection and segmentation difficult. Consequently, the development of reproducible, objective, and unbiased methods for medical image analysis has become a major research focus in recent years.

Algorithms for brain metastasis detection can be broadly divided into two categories: (*i*) classical and (*ii*) (deep) machine learning approaches [4, 17]. Ambrosini et al. [1] introduced 3D spherical tumor models for lesion detection in single-modal magnetic resonance images (MRIs), although such methods may fail to identify tumors with irregular morphology or atypical intensity characteristics. Farjam et al. [6] employed optimized 3D spherical shell templates for detecting small lesions, whereas Pérez-Ramírez et al. [18] proposed a template-matching framework for the same purpose. These techniques are frequently complemented by machine learning methods, including unsupervised clustering, to identify lesion candidates and reduce the number of false-positive detections through subsequent classification [19]. It is important to note that many of these early studies were conducted at a single institution and lacked pathological confirmation of metastatic lesions—it limits the generalizability of their findings [19]. In response to these limitations, the development of large-scale, objective, multi-center benchmarks has become increasingly important. One notable example is BraTS-METS [16, 15], established within the well-known Brain Tumor Segmentation (BraTS) family of challenges. BraTS-METS provides a standardized framework for evaluating emerging algorithms for brain metastasis detection and segmentation. In the 2025 edition of BraTS-METS, our winning solution [2] was based purely on nnU-Nets ("no-new U-Nets") [8]—a self-configuring deep learning model for medical image analysis. Of note, nnU-Nets have established the state of the art in a variety of medical image analysis segmentation tasks for different organs and image modalities [9]. The other BraTS-METS 2025 approaches included a combination of various deep learning architectures within the MONAI Auto3DSeg framework [11] and the approach benefiting from a combination of the Dice and Focal loss functions [20] in the training process. Finally, the federated benchmarking of (not only) medical artificial intelligence models has become increasingly important in the field, in order to objectify their thorough assessment and quantifying their generalization abilities [10, 22].

In this article, we build upon our last year's efforts, and extend our technique by building an ensemble combining two different versions of the U-Mamba model [14]. U-Mamba builds upon nnU-Nets, but additionally it is able to simultaneously capture local features and aggregate long-range dependencies thanks to benefiting from the advantages of fully-convolutional neural networks and state space sequence models [7, 5, 12]. In the experimental analysis, we verified several variants of the U-Mamba models and post-processing routines which led

us to developing an ensemble of two U-Mamba models, with one model specifically targeting small lesions, and utilizing the CATMIL loss [13].

The article is structured as follows. Section 2 discusses the 2026 BraTS-METS dataset used in this study. Our method is presented in Section 3. We report and discuss the experimental results in Section 4. Section 5 concludes the paper.

2 Data

This study exploits the BraTS-METS 2026 datasets, including the following multi-parametric magnetic resonance imaging (MRI) series:

- pre-contrast T1-weighted (T1W),
- post-contrast T1-weighted (T1C),
- T2-weighted (T2W),
- T2-weighted Fluid Attenuated Inversion Recovery (FLAIR).

The dataset is a retrospective compilation of pre- and post-treatment brain metastases mpMRI scans obtained from various institutions under standard clinical conditions. However, it includes the data captured in nine worldwide institutions, reflecting different equipment (thus image quality) and imaging protocols. The images were annotated using segmentation masks generated through the STAPLE-based fusion [21] of multiple brain metastasis segmentation algorithms. The resulting consensus labels were subsequently reviewed and manually refined by neuroradiology experts with varying levels of seniority and experience, following a standardized annotation protocol that was consistently applied throughout the entire ground-truth generation process. The final annotations were validated and approved by experienced board-certified neuroradiologists.

The ground-truth labels in the BraTS-METS 2026 are aligned with the previous year's edition, and included the following classes of interest:

- **Nonenhancing tumor core (NETC)**: all portions of the tumor core without contrast enhancement that are enclosed by enhancing tumor (ET). It represents the bulk of the tumor, being a part which is typically considered for surgical excision.
- **Surrounding non-enhancing FLAIR hyperintensity (SNFH)**: peritumoral edematous and infiltrated tissue, defined by the abnormal hyperintense signal envelope on the T2 FLAIR volumes, which includes the infiltrative non enhancing tumor, together with the vasogenic edema in the peritumoral region. The authors of the dataset emphasized that the non tumor related FLAIR signal abnormality, such as prior infarcts or microvascular ischemic white matter changes, are not included here.
- **Enhancing Tumor (ET)**: all tumor areas with noticeable contrast enhancement on postcontrast T1-weighted images. Similarly, the adjacent blood vessels, bleeding or intrinsic T1 hyperintensity are not included here.
- **Resection Cavity (RC)**: the resection of region within the brain tissue in the post-treatment cases.

In the 2026 BraTS-METS edition, the organizers added a detection leaderboard, with the intention of promoting algorithms sensitive in detecting lesions. The organizers emphasized that it is clinically relevant to cases of small ($< 27\,\mathrm{mm}^3$) lesions that either need to be counted as separate entities, or that they need to be quantified separately.

In total, there are 1296/179/303 MRI studies with ground-truth annotations included in the 2026 BraTS-METS training, validation, and test subsets. All results reported in this paper are obtained for the validation set (for which the participants do not have access to the ground-truth delineations, and all metrics are calculated by the official evaluation server), unless stated otherwise.

3 Methods

3.1 Detection and Segmentation of Brain Metastases

In this work, we expand our last year's BraTS-METS winning solution by incorporating the U-Mamba models into our framework. U-Mamba extends nnU-Net by simultaneously capturing local features and modeling long-range dependencies, leveraging the strengths of both fully-convolutional neural networks (automatically optimized using the nnU-Net framework, following the AutoML paradigm), and state space sequence models [7, 5, 12]. Additionally, in order to specialize a model in capturing small brain metastases, we exploit the recently-proposed CATMIL loss function [13], which jointly optimizes the voxel-level accuracy (important to precise segmentation) and lesion-level detection performance of the model. CATMIL combines (*i*) the Component-Adaptive Tversky (CAT) term into the loss function, together with (*ii*) a lesion-level Multiple Instance Learning (MIL) term. In (*i*), the contribution of individual lesion components are re-weighted, so that "small" structures have a stronger influence during training. In (*ii*), the model is encouraged to detect metastatic lesions. Overall, we investigated three separate (base) models (Table 1): our last-year winning solution (nnU-Net), U-Mamba without and with the CATMIL loss incorporated into the training process (in the no-CATMIL variant, we use the loss function which is the unweighted sum of the Dice loss and cross-entropy).

In our preliminary experimental study, we investigated the detection and segmentation performance of the separate (base) models over the training set within the BraTS-METS 2026 dataset. In Fig. 1, we present the relationship between the detection metrics (precision and recall, calculated for the predicted and ground-truth metastatic lesions), obtained using U-Mamba without and with utilizing CATMIL. In Fig. 2, on the other hand, we visualize the segmentation metrics for these two models (all metrics are reported for the training set of the BraTS-METS 2026 dataset, as only for this subset the ground-truth delineations are available for such analysis). It is evident that U-Mamba$^{\mathrm{CATMIL}}$ "specializes" in detecting small lesions, as it was expected. Here, we can hypothesize that it might be appropriate to utilize the predictions of U-Mamba for larger lesions, and those obtained using U-Mamba$^{\mathrm{CATMIL}}$—for the small(er) ones. To find an appropriate threshold τ, below which the U-Mamba$^{\mathrm{CATMIL}}$ predictions

Table 1. The methods investigated in this study. We **boldface** the submission identifier (Sub. ID) of the ensemble model which was dockerized and considered our final solution in the BraTS-METS 2026 challenge.

Sub. ID	Symbol	Description
		Separate (base) models
9769964	nnU-Net	Our BraTS-METS 2025 winning solution (nnU-Net) [2]
9770091	U-Mamba	U-Mamba [14]
9771408	U-Mamba$^{\text{CATMIL}}$	U-Mamba with the CATMIL [13] loss
		Ensemble models
9772708	$\mathcal{E}^{\tau_1}$	U-Mamba used for predictions $\geq$20 mm^3, Mamba$^{\text{CATMIL}}$ used for smaller predictions
9772917	$\mathcal{E}^{\tau_2}_{>20}$	U-Mamba used for predictions $\geq$12720 mm^3, Mamba$^{\text{CATMIL}}$ used for smaller predictions; all predictions $\leq$20 mm^3 are removed (with the hypothesis they may be false-positives)
9773219	$\mathcal{E}^{\tau_2}_{>20,\,\text{RC}}$	As in $\mathcal{E}^{\tau_2}_{>20}$, but also all RC class predictions from Mamba$^{\text{CATMIL}}$ were zeroed, and all U-Mamba's RC class predictions are inserted back in the final segmentation (except for the voxels with non-RC predictions, i.e., those which are not the background)

are used, we swept through the possible volume thresholds within the training subset of the 2026 BraTS-METS dataset (Fig. 3). Here, e.g., in (b), we additionally applied a threshold of 5 mm^3, below which we removed all predictions, hypothesizing that they are likely false positives. Finally, in (c), we searched for the threshold for removing such potential false positive detections (in the sweeps, we were maximizing all detection metrics). According to this analysis, we built three ensembles (Table 1), with $\tau \in \{\tau_1, \tau_2\}$, and $\tau_1 = 20\,\text{mm}^3$, and $\tau_1 = 12720\,\text{mm}^3$, and all lesions smaller than 20 mm^3 are treated as false positives in the ensembles (this threshold was selected experimentally, and is smaller than the threshold of 27 mm^3 for calculating the segmentation metrics imposed by the BraTS-METS 2026 organizers), as discussed in Table 1. Finally, we observed that U-Mamba$^{\text{CATMIL}}$ failed to learn the RC class. Therefore, to account for this, we remove all RC predictions delivered by U-Mamba$^{\text{CATMIL}}$ in post-processing, and inject the U-Mamba's RC predictions in the final segmentation mask (only for those pixels which were not assigned to any other class).

3.2 Evaluation Criteria

In BraTS-METS 2026, for segmentation performance evaluation, the following subject-wise metrics are used:

- **Dice Similarity Coefficient (DSC)**, a widely adopted metric for quantifying the overlap between predicted and reference segmentation masks. This metric should be maximized, with one indicating the perfect score.
- **Normalized Surface Distance (NSD)**, which incorporates a predefined tolerance parameter and provides complementary information to overlap-based metrics by evaluating the accuracy of predicted boundaries. This metric should be maximized, with one indicating the perfect score.

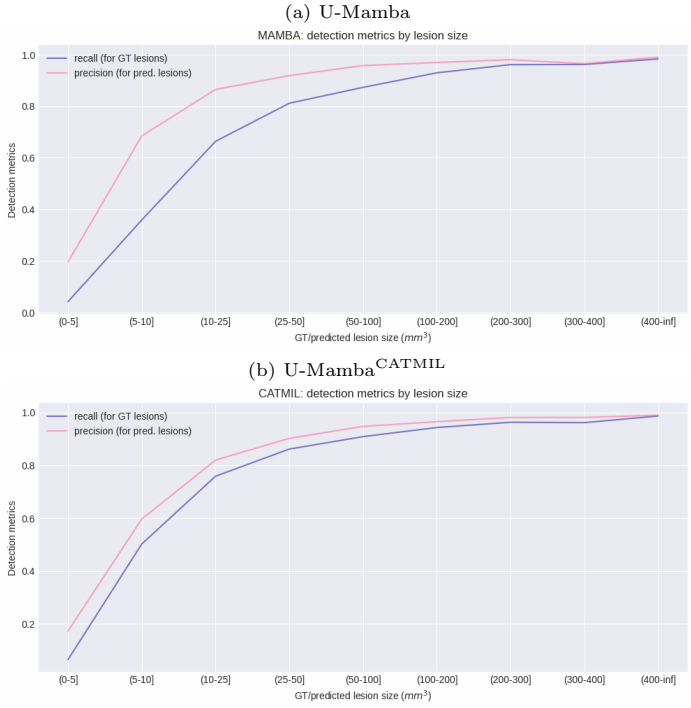

Fig. 1. The detection metrics (precision and recall) obtained using U-Mamba and U-Mamba$^{\text{CATMIL}}$, with respect to the ground-truth/predicted lesion size (note the different range on the Y axis in both plots). It is observable that—as expected—U-Mamba$^{\text{CATMIL}}$ delivers higher-quality detection performance for small lesions.

For metastatic lesion detection evaluation, the lesion-wise F1 score metric is used, and it is defined as the harmonic mean of precision and recall. It assesses the balance between correctly detected lesions and false positive/negative predictions, providing insight into potential over- or under-segmentation tendencies. This metric should be maximized, with one indicating the perfect performance.

The organizers indicated that the segmentation-based evaluation metrics are computed only for lesions with a volume larger than $27\,\text{mm}^3$.

4 Experiments and Discussion

4.1 Experimental Setup

All the experiments were executed using a computational server equipped with an NVIDIA A100-PCIE-40GB Graphics Processing Unit (GPU). All nnU-Net and U-Mamba hyperparameters were kept as suggested in their default values[3].

[3] The nnU-Net parameterization used in our study can be found in the nnU-Net original repository available at: https://github.com/MIC-

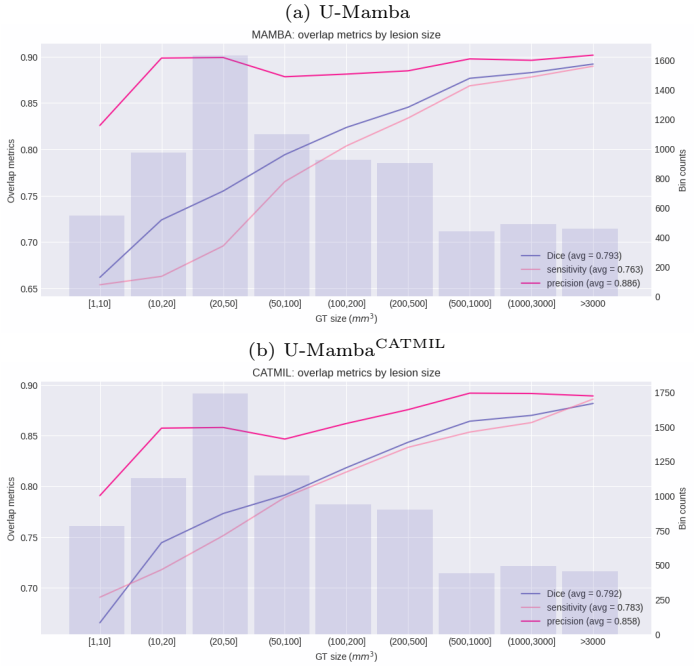

Fig. 2. The segmentation overlap (voxel-wise) metrics (Dice, sensitivity, precision) obtained using U-Mamba and U-Mamba$^{\text{CATMIL}}$ over the training set of BraTS-METS 2026, with respect to the ground-truth size (note the different range on the Y axes).

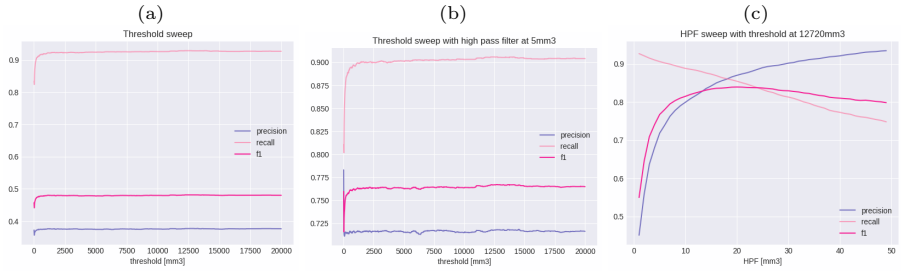

Fig. 3. Searching for an appropriate threshold, below which U-Mamba$^{\text{CATMIL}}$ predictions are used (above this threshold, U-Mamba predictions are used, hypothesizing that U-Mamba$^{\text{CATMIL}}$ is specialized in detecting and segmenting small lesions, thanks to the utilization of the CATMIL loss function). In (a), we do not apply any additional threshold for removing small objects, whereas in (b) we remove the lesions smaller than $5\,\text{mm}^3$. In (c), we search for the best "high-pass filtering" (HPF) threshold (to remove small objects, with a hypothesis that they are likely false positive detections), with the U-Mamba/U-Mamba$^{\text{CATMIL}}$ threshold set at $12720\,\text{mm}^3$. These experiments were performed over the training subset of the 2026 BraTS-METS dataset.

DKFZ/nnUNet/blob/master/nnunetv2/training/nnUNetTrainer/nnUNetTrainer.py (last access: July 26, 2026), whereas the U-Mamba implementation, together with its parameterization, used in this work is available at: https://github.com/bowang-lab/U-Mamba. The implementation of our ensemble approach is available at: https://bitbucket.org/fphealthcare/brats2026-ensemble.

Training the ensemble selected for evaluation took approx. 172 h, whereas its inference was approx. 4 s per one multi-modal MRI scan on average. To ensure reproducibility and traceability of the experimentation, we exploited MLFlow [3]. Our team identifier on the Synapse platform is **GLI2026**.

4.2 Results and Discussion

The experimental results, as returned by the evaluation server for the validation set of 2026 BraTS-METS are gathered in Tables 2–3 (for the segmentation and detection metrics, respectively). Based on the results, we picked the $\mathcal{E}^{\tau_2}_{>20,\,\mathrm{RC}}$ ensemble, as the one which balances the segmentation and detection metrics. The ensemble method delivers the results which are statistically significantly better than our last year's nnU-Nets, according to the Wilcoxon matched-pairs signed rank test at $p < 0.05$ calculated for the global DSC over the validation set for ET and TC, with the p values of 0.019 and 0.005 for ET and TC, and for the lesionwise DSC for ET, TC and WT with the p values of 0.003, 0.001, and 0.015, respectively. It might be observed that U-Mamba$^{\mathrm{CATMIL}}$ outperforms the other methods in detecting metastatic lesions. However, the segmentation metrics obtained for this model were substantially worse when compared to U-Mamba—this was the motivation behind bundling these two models into an ensemble. In Figs. 4–6, we present visual examples (together with the quality metrics, including the overlap metrics: Dice, precision and recall, and lesion-wise metrics: precision and recall) of predictions obtained using the final ensemble model for "large", "medium" and "small" metastatic lesions from the 2026 BraTS-METS training set (note that the mapping of lesion sizes to these "classes" was subjective). The models failed to spot small instances of RC—this class examples are under-represented (and likely heterogeneous), and building expert models for this specific class might potentially be a step toward addressing this issue.

Table 2. The segmentation results, as returned by the evaluation server for the 2026 BraTS-METS validation set. The best results are **boldfaced**.

	ET		RC		TC		WT	
Method	DSC	NSD	DSC	NSD	DSC	NSD	DSC	NSD
nnU-Net	0.667	0.726	0.500	0.384	0.694	0.740	0.648	0.668
U-Mamba	0.679	0.741	**0.503**	0.426	0.714	0.766	0.675	0.700
U-Mamba$^{\mathrm{CATMIL}}$	0.684	0.743	0.139	0.122	0.708	0.756	0.616	0.633
$\mathcal{E}^{\tau_1}$	0.682	0.740	0.141	0.124	0.718	0.766	0.646	0.669
$\mathcal{E}^{\tau_2}_{>20}$	**0.708**	**0.767**	0.300	0.251	**0.740**	**0.789**	**0.694**	**0.711**
$\mathcal{E}^{\tau_2}_{>20,\,\mathrm{RC}}$	**0.708**	**0.767**	0.499	**0.427**	**0.740**	**0.789**	**0.694**	**0.711**

5 Conclusion

In this work, we tackled the problem of automatic the process of detecting and segmenting brain metastatic lesions in MRI scans. These steps are critically

Table 3. The "*small instance*" detection results (F1 score), as returned by the evaluation server for the 2026 BraTS-METS validation set. The best results are **boldfaced**.

Method	ET	TC	WT	RC
nnU-Net	0.433	0.440	0.358	0.000
U-Mamba	0.474	0.483	0.381	0.000
U-Mamba$^{\text{CATMIL}}$	**0.500**	**0.508**	**0.397**	0.000
$\mathcal{E}^{\tau_1}$	0.211	0.214	0.135	0.000
$\mathcal{E}^{\tau_2}_{>20}$	0.450	0.457	0.357	0.000
$\mathcal{E}^{\tau_2}_{>20,\,\text{RC}}$	0.450	0.457	0.357	0.000

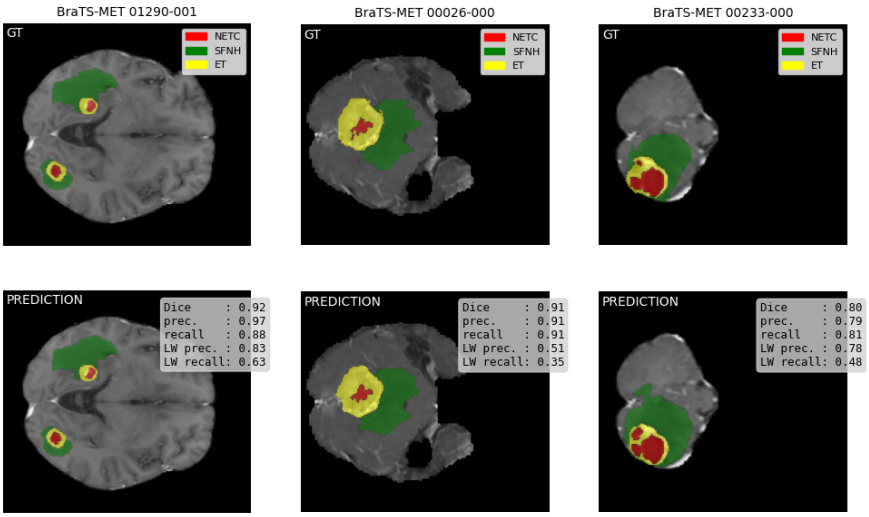

Fig. 4. "Large"-sized lesions—examples of predictions, together with the quality metrics, obtained using our method for MRIs from the training subset of 2026 BraTS-METS. The "large" metastatic lesions were selected subjectively.

important to objectively monitor the disease. We built upon our framework developed for BraTS-METS 2025, which exploited nnU-Nets and was the winning solution. In this study, we utilized U-Mamba models that benefit from automatically-optimized fully-convolutional neural networks and state space sequence models, and bundled two "specialized" U-Mamba models into an ensemble. The experimental results over the training and validation subsets of the 2026 BraTS-METS dataset showed that this ensemble effectively balances the detection and segmentation quality, outperforming the other investigated methods, including our last year's nnU-Nets.

Acknowledgments. The project is co-financed from the Medical Research Agency and European Union funds. The funds come from the National Recovery and Resilience Plan (NRRP) and the European Union's NextGeneration EU (NGEU) financial instrument (2024/ABM/06/KPO/KPOD.07.07-IW.07-0295/24-00). JN was supported

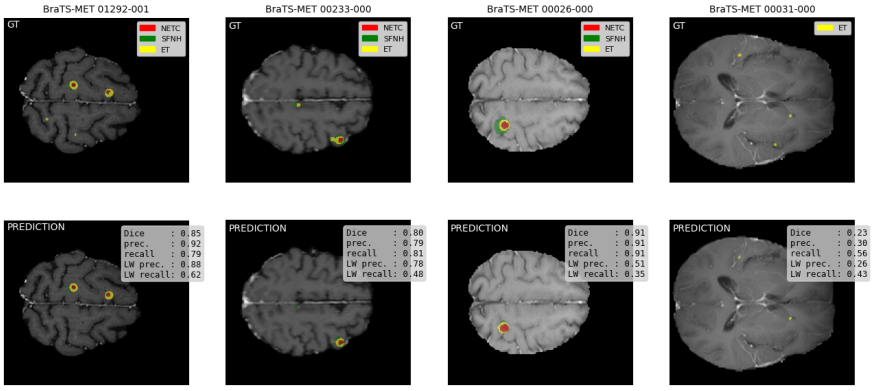

Fig. 5. "Medium"-sized lesions—examples of predictions, together with the quality metrics, obtained using our method for MRIs from the training subset of 2026 BraTS-METS. The "medium" metastatic lesions were selected subjectively.

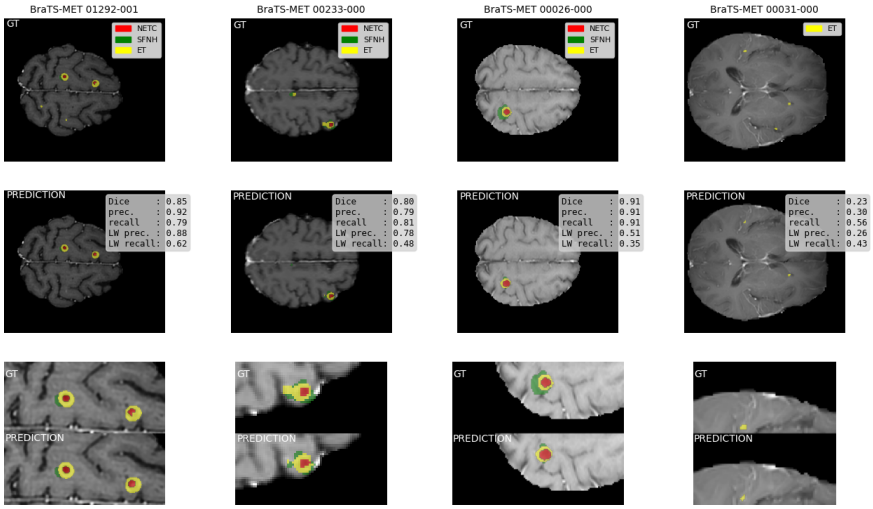

Fig. 6. "Small"-sized lesions—examples of predictions, together with the quality metrics, obtained using our method for MRIs from the training subset of 2026 BraTS-METS. The "small" metastatic lesions were selected subjectively.

by Silesian Univ. of Tech. grant for maintaining and developing research potential. The authors thank Daria Bernys (Graylight Imaging) for helping us manage this study.

This paper is in memory of Dr. Grzegorz Nalepa, an extraordinary scientist, pediatric hematologist/oncologist, and a compassionate champion for kids at Riley Hospital for Children, Indianapolis, USA, who helped countless patients and their families through some of the most challenging moments of their lives.

Disclosure of Interests. The authors have no competing interests.

References

1. Ambrosini, R.D., Wang, P., O'Dell, W.G.: Computer-aided detection of metastatic brain tumors using automated three-dimensional template matching. Journal of Magnetic Resonance Imaging **31**(1), 85–93 (2010). https://doi.org/https://doi.org/10.1002/jmri.22009
2. Bancerek, M., Rudzki, P., Nalepa, J.: Segmentation of Pre and Post-treatment Brain Metastases Using nnU-Nets. In: Bakas, S., Dennis, E., Astaraki, M., Baid, U., Conte, G.M., Foltyn-Dumitru, M., Jiang, Z., Labella, D., Metz, M.C., Anazodo, U., de Verdier, M.C., Kofler, F., Li, H.B., Linguraru, M.G., Maleki, N. (eds.) Segmentation, Classification, and Synthesis for Brain Tumors and Traumatic Brain Injuries. pp. 161–172. Springer Nature Switzerland, Cham (2026)
3. Chen, A., Chow, A., Davidson, A., et al.: Developments in MLflow: A System to Accelerate the Machine Learning Lifecycle. In: Proceedings of the Fourth International Workshop on Data Management for End-to-End Machine Learning. DEEM '20, Association for Computing Machinery, New York, NY, USA (2020). https://doi.org/10.1145/3399579.3399867, https://doi.org/10.1145/3399579.3399867
4. Cho, S.J., Sunwoo, L., Baik, S.H., Bae, Y.J., Choi, B.S., Kim, J.H.: Brain metastasis detection using machine learning: a systematic review and meta-analysis. Neuro-Oncology **23**(2), 214–225 (10 2020). https://doi.org/10.1093/neuonc/noaa232
5. Dao, T., Gu, A.: Transformers are SSMs: Generalized models and efficient algorithms through structured state space duality. In: International Conference on Machine Learning (ICML) (2024)
6. Farjam, R., Parmar, H.A., Noll, D.C., Tsien, C.I., Cao, Y.: An approach for computer-aided detection of brain metastases in post-Gd T1-W MRI. Magnetic Resonance Imaging **30**(6), 824–836 (2012). https://doi.org/https://doi.org/10.1016/j.mri.2012.02.024
7. Gu, A., Dao, T.: Mamba: Linear-time sequence modeling with selective state spaces. arXiv preprint arXiv:2312.00752 (2023)
8. Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods **18**(2), 203–211 (Feb 2021). https://doi.org/10.1038/s41592-020-01008-z
9. Isensee, F., Wald, T., Ulrich, C., Baumgartner, M., Roy, S., Maier-Hein, K., Jäger, P.F.: nnU-Net Revisited: A Call for Rigorous Validation in 3D Medical Image Segmentation . In: proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. vol. LNCS 15009. Springer Nature Switzerland (October 2024)
10. Karargyris, A., Umeton, R., Sheller, M.J., et al.: Federated benchmarking of medical artificial intelligence with MedPerf. Nature Machine Intelligence **5**(7), 799–810 (Jul 2023). https://doi.org/10.1038/s42256-023-00652-2, https://doi.org/10.1038/s42256-023-00652-2
11. Krikorian, W., Purwar, A.: Automated Segmentation for the Brain Tumor Segmentation (BraTS) Metastases 2025 Challenge Using Multi-Architectural Deep Learning. In: Bakas, S., Dennis, E., Astaraki, M., Baid, U., Conte, G.M., Foltyn-Dumitru, M., Jiang, Z., Labella, D., Metz, M.C., Anazodo, U., de Verdier, M.C., Kofler, F., Li, H.B., Linguraru, M.G., Maleki, N. (eds.) Segmentation, Classification, and Synthesis for Brain Tumors and Traumatic Brain Injuries. pp. 173–181. Springer Nature Switzerland, Cham (2026)

12. Lahoti, A., Li, K.Y., Chen, B., Wang, C., Bick, A., Kolter, J.Z., Dao, T., Gu, A.: Mamba-3: Improved sequence modeling using state space principles (2026), https://arxiv.org/abs/2603.15569
13. Luu, M.S.K., Pavlovskiy, E.N., Tuchinov, B.N.: Component-Adaptive and Lesion-Level Supervision for Improved Small Structure Segmentation in Brain MRI (2026), https://arxiv.org/abs/2604.08015
14. Ma, J., Li, F., Wang, B.: U-Mamba: Enhancing Long-range Dependency for Biomedical Image Segmentation. arXiv preprint arXiv:2401.04722 (2024)
15. Maleki, N., Amiruddin, R., Moawad, A.W., et al.: Analysis of the MICCAI Brain Tumor Segmentation – Metastases (BraTS-METS) 2025 Lighthouse Challenge: Brain Metastasis Segmentation on Pre- and Post-treatment MRI (2025), https://arxiv.org/abs/2504.12527
16. Moawad, A.W., Janas, A., Baid, U., et al.: The Brain Tumor Segmentation (BraTS-METS) Challenge 2023: Brain Metastasis Segmentation on Pre-treatment MRI (2024), https://arxiv.org/abs/2306.00838
17. Park, Y.W., Lee, N., Ahn, S.S., Chang, J.H., Lee, S.K.: Radiomics and deep learning in brain metastases: Current trends and roadmap to future applications. Investig Magn Reson Imaging **25**(4), 266–280 (Dec 2021). https://doi.org/10.13104/imri.2021.25.4.266
18. Perez-Ramírez, U., Arana, E., Moratal, D.: Brain metastases detection on MR by means of three-dimensional tumor-appearance template matching. Journal of Magnetic Resonance Imaging **44**(3), 642–652 (2016). https://doi.org/https://doi.org/10.1002/jmri.25207
19. Sunwoo, L., Kim, Y.J., Choi, S.H., Kim, K.G., Kang, J.H., Kang, Y., Bae, Y.J., Yoo, R.E., Kim, J., Lee, K.J., Lee, S.H., Choi, B.S., Jung, C., Sohn, C.H., Kim, J.H.: Computer-aided detection of brain metastasis on 3D MR imaging: Observer performance study. PLOS ONE **12**(6), 1–18 (06 2017). https://doi.org/10.1371/journal.pone.0178265
20. Umeh, F., Yordanov, N.Y., Maleki, N., Amiruddin, R., Moawad, A., Pytlarz, M., Chukwurah, C., Aboian, M.: Taking Advantage of MONAI and DiNTS Frameworks to Develop a State-of-the-Art Algorithm for Automatic Segmentation of Brain Metastases. In: Bakas, S., Dennis, E., Astaraki, M., Baid, U., Conte, G.M., Foltyn-Dumitru, M., Jiang, Z., Labella, D., Metz, M.C., Anazodo, U., de Verdier, M.C., Kofler, F., Li, H.B., Linguraru, M.G., Maleki, N. (eds.) Segmentation, Classification, and Synthesis for Brain Tumors and Traumatic Brain Injuries. pp. 182–189. Springer Nature Switzerland, Cham (2026)
21. Warfield, S., Zou, K., Wells, W.: Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Transactions on Medical Imaging **23**(7), 903–921 (2004). https://doi.org/10.1109/TMI.2004.828354
22. Zenk, M., Baid, U., Pati, S., et al.: Towards fair decentralized benchmarking of healthcare AI algorithms with the Federated Tumor Segmentation (FeTS) challenge. Nature Communications **16**(1), 6274 (Jul 2025). https://doi.org/10.1038/s41467-025-60466-1, https://doi.org/10.1038/s41467-025-60466-1