Supplementary Material Anatomy-Preserving Rare-Subtype Adaptation for Pediatric Brain Tumor Segmentation

Hari Mohan Rai[1] and Dauren Omarbekov[1]

Nazarbayev University, Astana, Kazakhstan
hari.rai@nu.edu.kz, dauren.omarbekov@nu.edu.kz

Equation numbers below refer to the main paper: Eq. (1) is the region decoding, Eq. (2) the proposal set Ω, and Eq. (3) the budgeted assembly. Code, container recipe and the full verification record are at `https://github.com/1nqi/hmnunet-brats2026-task2`.

S1 Per-Case Inference Procedure

Figure S1 and Algorithm S1 give the same pipeline as a schematic and as pseudocode. The two are deliberately not redundant: the figure shows which trained configuration sits in which ensemble and that the hybrid ensemble drives both the region of interest and the budget test, while the pseudocode gives the control flow, including the empty tumor-core skip and the order of the two guards.

Algorithm S1 Docker inference for one case

Require: Four co-registered MRI volumes
Predict region probabilities with the primary ensemble (three configurations $\times$ five folds, test-time augmentation)
Average in single precision, decode with Eq. (1), post-process once $\Rightarrow P$
Repeat with the hybrid ensemble $\Rightarrow$ raw H_{raw} and post-processed H
if tumor core of H_{raw} is empty **then**
    $\hat{Y} \leftarrow P$ ▷ specialist stage skipped
else
    Crop the tumor-core bounding box with a 16-voxel margin; build the six-channel specialist input
    Predict with the five-fold ET specialist; paste back under identity geometry
    Fuse NETC$\rightarrow$specialist inside H_{raw}; post-process once $\Rightarrow F$
    Form Ω by Eq. (2) and apply Eq. (3)
end if
if tumor core of $\hat{Y}$ is empty and tumor core of H is not **then**
    $\hat{Y} \leftarrow H$ ▷ terminal guard
end if
Assert the invariants, restore image geometry, and write an unsigned 8-bit NIfTI volume atomically
Ensure: labels are in $\{0, 1, 2, 3\}$; TC, WT and CC equal those of P outside the guard

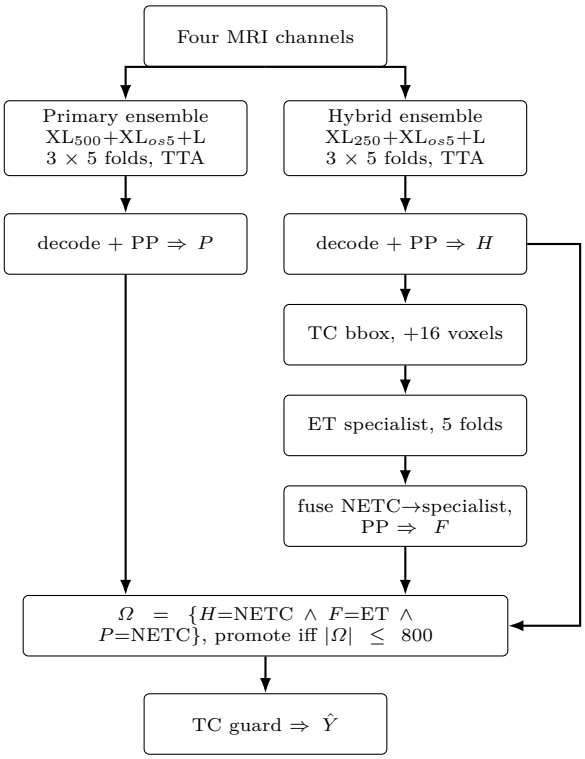

Fig. S1. Frozen Docker inference pipeline. The primary ensemble defines the submitted segmentation; the hybrid ensemble localizes the tumor-core region of interest for the specialist and is the fallback of the terminal guard. The final stage can only relabel NETC as ET, is rejected in full above an inclusive 800-voxel budget, and ends with a tumor-core guard. Background, ED, TC, WT, and the cystic mask are preserved exactly.

S2 Container Construction and Build Gates

The submitted image is built on a frozen parent image and installs no additional packages. Its build-time gates assert the locked versions of every scientific dependency (Python 3.11.11, nnU-Net 2.8.1, PyTorch 2.6.0+cu124, NumPy 2.2.2, SimpleITK 2.5.5, acvl-utils 0.2.6), the SHA-256 hashes of the three trainer implementations, an exact 21-file manifest for each active model family, an exact union of 25 checkpoints across the two ensembles and the specialist, and the absence of validation, probability or oracle artifacts anywhere inside the image. The inference operators are imported from a single hash-locked module and the build fails if the runner defines them a second time. Both the parent image and the submitted image are referenced by immutable content digests, so the container that was verified cannot be silently replaced by a rebuild.

S3 Synthetic Boundary Tests

A suite of tests runs inside the final image and requires no data. It pins the strict decision boundary at 0.5, the paint order of Eq. (1), the removal of edema, the connected-component thresholds at 99 versus 100 voxels for ET and 49 versus 50 for NETC, the cystic retiering at 9, 10, 49 and 50 voxels, the inclusive budget boundary at 799, 800 and 801 voxels, the truth table of the terminal guard, and the exact construction of the six-channel specialist crop. Any drift in a threshold therefore fails the build rather than changing a score silently.

S4 Replay Measurements

Section 3.5 of the main paper reports the replay of the submitted image against the frozen reference: ten of 91 cases differed, by twelve voxels in total, with presence, connected-component counts and instance topology unchanged in every region and a minimum per-case agreement Dice of 0.99960. Two further measurements support it.

The primary ensemble alone reproduces its own reference to five voxels across four cases under the same protocol. Comparing two independent full replays against each other gives a third and consistent measurement: once the voxels the transplant accounts for are set aside, the two replays differ by eight voxels in eight cases. That comparison also audits the transplant from outside the container, since the number of NETC→ET voxels measured directly from the two mask sets equals the proposal size the container reports for every admitted case, and is exactly zero for every rejected case.

S5 Container Compliance

Table S1 lists the resource usage measured during the full 91-case replay against the Task 2 evaluation limits. Host memory is the tightest of them, because the

pipeline retains the raw segmentations of the cohort while the specialist stage runs; GPU memory, by contrast, is used at roughly a third of the available budget.

Table S1. Container compliance against the Task 2 evaluation limits, as measured during the full 91-case replay.

Resource or contract	Challenge limit	Verified submission
GPU memory	NVIDIA A10G, 24 GB	8.3 GiB peak
Wall-clock time	12 h	89.6 min for 91 cases
Host memory	48 GiB	44.0 GiB peak
Shared memory	16 GiB	within limit
Storage	200 GB	within limit
Network	disabled	no runtime network access
Input	read-only `/input`	no input writes
Output	flat `/output`	91 valid NIfTI files

S6 Qualitative Examples

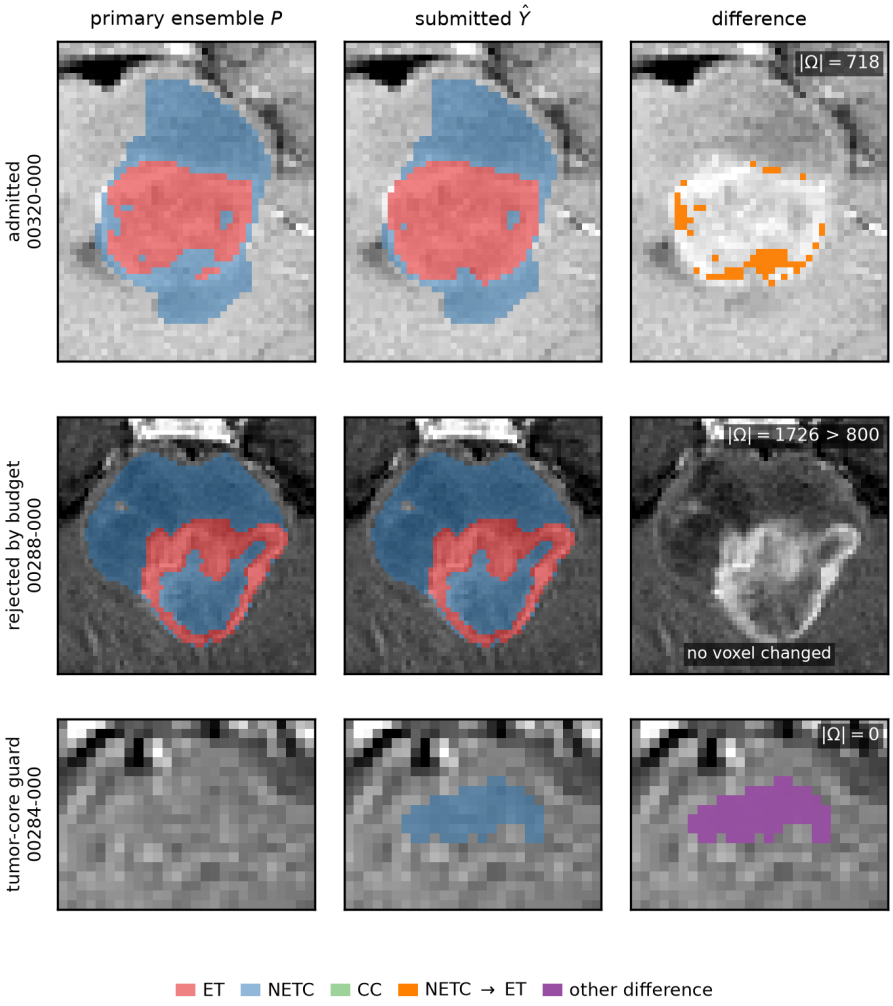

Fig. S2. One case per regime on the T1C image: the primary ensemble P, the submitted segmentation $\hat{Y}$, and their difference. Top, an admitted proposal. Middle, a proposal above the budget, rolled back with no voxel changed. Bottom, the terminal guard. No ground truth exists for this cohort, so the panels show the mechanism, not correctness.