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.

# Supplementary Material

Agamdeep S. Chopra[1] and Mehmet Kurt[1]

Department of Mechanical Engineering, University of Washington, Seattle, WA, USA

**Partial Information Decomposition as a Multi-Contrast 3D MRI Selection Strategy for Resource-Constrained Deep Neural Network Training in Brain Tumor Segmentation**

## Appendix

### Comparison with Simpler Selection Criteria

We compared the proposed selector with simpler criteria computed from the same training-only quantized representations used for PID. The comparison includes weighted joint MI, top-two individual MI, sequential mRMR-MID, and uniform random pair selection. Associated segmentation performance was obtained from the already evaluated pair models; test labels were not used by any selector. Clinical face-validity and post-hoc Shapley are included as contextual audits rather than equivalent pre-training algorithms.

**Table 1.** Comparison of training-only selection criteria using the same quantized representations. Associated Dice is the test performance of the corresponding already evaluated pair model.

| Criterion | Selected pair / outcome | Test Dice |
|---|---|---|
| Proposed MMI-PID ($\lambda = 0.5$) | T1c+T2-FLAIR | 0.676 |
| Weighted joint MI | T1c+T2-FLAIR | 0.676 |
| Sequential mRMR-MID | T2-FLAIR+T1c | 0.676 |
| Top-two individual MI | T2-FLAIR+T2w | 0.500 |
| Uniform random pair | Expectation over six pairs | 0.559 |
| Clinical face-validity | T1c+T2-FLAIR | 0.676 |
| Post-hoc Shapley | T2-FLAIR+T1c | 0.676 |

The $\lambda$ sensitivity can be evaluated exactly from the PID terms without re-estimating representations. The score defined in Eq. (2) of the main manuscript is affine in $\lambda$. For T1c+T2-FLAIR, the weighted joint-MI term was 0.702 and the weighted non-redundant term was 0.521; both were the largest among all six pairs. Consequently, T1c+T2-FLAIR remains rank 1 for every $\lambda \geq 0$, including $\lambda = 0$ (joint MI alone). This establishes robustness to the score-weight parameter for the evaluated representation, but not to $K$, latent dimension, foreground-sampling fraction, or representation-training seeds.

### Complete 11-Model Test Results

Table 2 provides the complete patient-level macro-averaged test metrics for all 11 evaluated input configurations. These are the same results used for the statistical comparisons in the main paper.

**Table 2.** Complete test-set performance for all 11 input configurations. The PID-selected pair is marked with an asterisk.

| Inputs | Ch. | HD95 | Avg. Dice | Sens. | Prec. |
|---|---|---|---|---|---|
| All four | 4 | **16.09** | **0.687** | 0.709 | **0.714** |
| T1c+T2-FLAIR* | 2 | 16.77 | 0.676 | **0.709** | 0.699 |
| T1c+T2w | 2 | 18.06 | 0.650 | 0.670 | 0.683 |
| T1c | 1 | 21.71 | 0.589 | 0.606 | 0.622 |
| T1n+T1c | 2 | 18.21 | 0.587 | 0.594 | 0.612 |
| T2w+T2-FLAIR | 2 | 20.93 | 0.500 | 0.572 | 0.511 |
| T1n+T2-FLAIR | 2 | 21.90 | 0.483 | 0.575 | 0.495 |
| T1n+T2w | 2 | 29.26 | 0.457 | 0.561 | 0.470 |
| T2-FLAIR | 1 | 21.40 | 0.457 | 0.539 | 0.469 |
| T2w | 1 | 31.23 | 0.393 | 0.473 | 0.410 |
| T1n | 1 | 24.90 | 0.377 | 0.434 | 0.424 |