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

Agamdeep S. Chopra[1][0009-0003-4386-7894] and Mehmet Kurt[1][0000-0002-5618-0296]

Department of Mechanical Engineering, University of Washington, Seattle, WA, USA
{achopra4,mkurt}@uw.edu

Abstract. Multi-contrast 3D MRI segmentation can be computationally demanding when all available sequences are used. We evaluate a pre-training Partial Information Decomposition framework that ranks input pairs according to their redundant, unique, and synergistic information about regional tumor burden and selects the highest-ranked pair for downstream training. Applied to T1n, T1c, T2w, and T2-FLAIR MRI, the framework selected T1c+T2-FLAIR. We then trained eleven architecturally identical lightweight 3D U-Nets using different input configurations. On an independent test cohort, T1c+T2-FLAIR was the strongest two-input configuration and ranked second overall in mean Dice (0.676 versus 0.687 for all four inputs). A separate Shapley analysis of the full-input model also identified T2-FLAIR and T1c as the most influential inputs and their pairwise interaction as the strongest. These findings support the practical value of PID-based pre-training selection for identifying compact, informative MRI input sets before costly 3D model development, while the present study remains a focused proof-of-concept.

Keywords: Partial information decomposition · MRI sequence selection · Brain tumor segmentation · Resource-constrained learning

1 Introduction

Multi-contrast MRI supports brain tumor segmentation because each sequence emphasizes complementary tissue characteristics. T1-weighted MRI depicts anatomy, contrast-enhanced T1 highlights enhancing tumor, T2 MRI is sensitive to fluid, and T2-FLAIR suppresses cerebrospinal fluid to emphasize edema and infiltrative abnormalities. Consequently, brain tumor benchmarks typically combine these sequences to support multi-region segmentation [13, 6]. High-performing 3D segmentation frameworks likewise assume that all available contrasts can be ingested jointly [4, 9].

This approach can be expensive in computationally constrained settings. Each additional sequence increases storage, preprocessing, data-transfer, memory, and training costs, which are amplified in 3D by large patch and volume tensors and by repeated experiments across folds, architectures, and hyperparameter settings. The practical question is therefore whether a smaller input subset can be selected before substantial resources are committed to downstream model development.

Standard subset selection relies on either exhaustive model training or single-source relevance measures. Exhaustive evaluation provides direct task performance estimates but scales poorly and must be repeated for each downstream architecture. Mutual-information-based methods offer a cheaper alternative by balancing target relevance against redundancy [14], but they may overlook synergy, whereby two sequences are jointly informative even when neither is individually dominant.

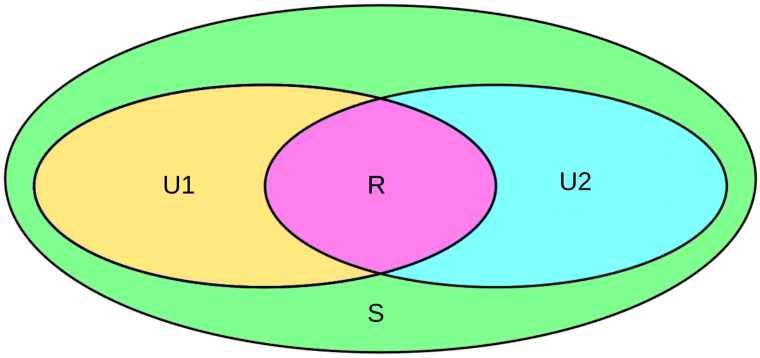

Fig. 1. Diagram representing partial information decomposition for two sources, X_1 and X_2, relative to a target Y. The total information is partitioned into source-specific contributions, U_1 and U_2, redundant information R provided by both sources, and synergistic information S available only from their joint observation.

Partial information decomposition (PID) separates target information into redundant, unique, and synergistic components [19, 2]. It has been used to define feature relevance and guide feature selection [20, 18], while recent work such as SFL-Net applied PID-guided latent factorization to disentangle shared, modality-specific, and cross-modality information in multi-contrast MRI-to-PET synthesis [3]. Building on this work, prior studies of MRI sequence contributions in brain tumor segmentation [21, 16], and contrast-level Shapley attribution [15], we introduce a general n-to-2 pre-training framework for multi-contrast MRI tumor segmentation. A compact autoencoder is trained for each input sequence, the resulting embeddings are quantized, and all $\binom{n}{2}$ pairs are ranked using a region-aware minimum mutual information PID (MMI-PID) score [1]. The highest-scoring pair is then selected for downstream model development.

We evaluate the framework using four CoRe-BT MRI sequences [6], comparing the full four-input model, all six two-input combinations, and four single-input models with patient-level statistical testing and Shapley attribution under a restricted compute budget. The four-input setting is intentionally tractable so that every candidate pair can be trained afterward as an audit of the pre-training selection.

2 Method

2.1 PID Pair Selection

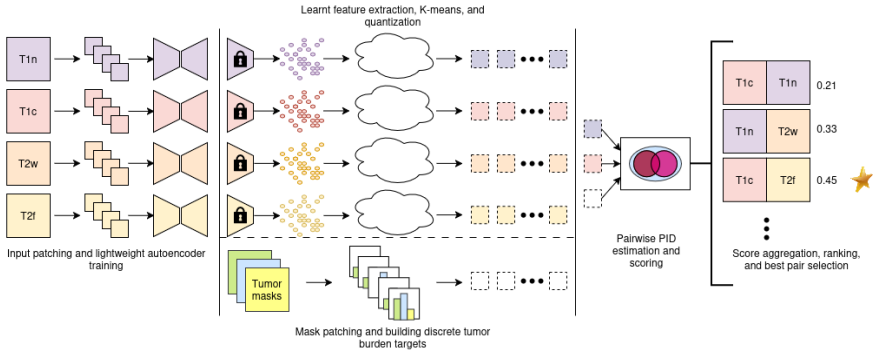

Fig. 2. PID pair-selection pipeline. Spatially aligned patches from each MRI sequence are encoded using sequence-specific autoencoders. The frozen latent representations are quantized into discrete source codes, while aligned tumor masks provide discretized regional tumor-burden targets. All two-input combinations are evaluated using region-aware PID, ranked by their aggregate scores, and the highest-ranked pair is retained for downstream model training.

Figure 2 summarizes the selection process. For each training subject, we sampled 128 spatially corresponding $32 \times 32 \times 32$ patches from all MRI sequences and the tumor mask. Seventy percent of patch centers were drawn from the tumor foreground. For each tumor region $c \in \{\text{edema}, \text{enhancing tumor}, \text{necrotic}\}$, tumor burden was defined as the fraction of patch voxels assigned to that region. Zero-burden patches formed class 0, while positive burdens were divided into three training-set quantile classes.

A separate shallow 3D autoencoder was trained for each MRI sequence to reconstruct its input patches using mean-squared error. The encoder comprised three $3 \times 3 \times 3$ convolutional blocks with stride two and 8, 16, and 32 output channels, respectively. Each convolution was followed by instance normalization and a ReLU activation. For a $32 \times 32 \times 32$ input patch, the encoder produced a $32 \times 4 \times 4 \times 4$ feature map, which was flattened and linearly projected to a

32-dimensional latent vector. After training, the encoders were frozen and the decoders were discarded. For each sequence m, a separate K-means model with $K = 16$ was fitted to the training latent vectors, and each cluster assignment defined the discrete source code q_m [7, 10]. Autoencoder optimization hyperparameters beyond the reconstruction objective were not part of the verified experiment record and are therefore not reported.

For every input pair (a, b) and tumor region c, mutual information was estimated from the empirical joint distribution of the two source codes and the discretized tumor-burden target T_c [5]. Following the bivariate PID framework [19], we decomposed the information that each pair of quantized source codes provided about the regional tumor-burden target. We used the MMI formulation as a computationally simple proxy for bivariate PID [1]. For input pair (a, b) and tumor region c, the PID terms were

$$\begin{aligned} R^c_{ab} &= \min\{I(q_a; T_c), I(q_b; T_c)\}, \\ U^c_{a|b} &= I(q_a; T_c) - R^c_{ab}, \qquad U^c_{b|a} = I(q_b; T_c) - R^c_{ab}, \\ S^c_{ab} &= I((q_a, q_b); T_c) - R^c_{ab} - U^c_{a|b} - U^c_{b|a}. \end{aligned} \tag{1}$$

To rank the candidate pairs, we defined an aggregate score inspired by PID-based feature-selection criteria that favor informative and non-redundant feature sets [20, 18]:

$$\mathrm{Score}(a, b) = \sum_c w_c \left[I((q_a, q_b); T_c) + \lambda \left(U^c_{a|b} + U^c_{b|a} + S^c_{ab} \right) \right], \qquad \lambda = 0.5, \tag{2}$$

where w_c was the normalized inverse frequency of positive patches for region c. Thus, joint target information determined the baseline pair score, while unique and synergistic information received an additional weight relative to redundant information. The selected pair was

$$\mathcal{M}^* = \arg \max_{\{a,b\} \subset \mathcal{M}} \mathrm{Score}(a, b), \qquad |\mathcal{M}^*| = 2. \tag{3}$$

All selection steps used training data only. Patient-level bootstrap resampling retained all patches from each sampled subject.

2.2 Lightweight 3D U-Net

The downstream segmentation model was a lightweight 3D U-Net with three resolution levels [4]. Each encoder and decoder block contained two $3 \times 3 \times 3$ convolutions followed by ReLU activations and instance normalization. Max pooling was used for downsampling, while nearest-neighbor interpolation followed by $1 \times 1 \times 1$ convolution was used for upsampling. A final $1 \times 1 \times 1$ convolution produced the output logits.

Models were trained using $64 \times 64 \times 64$ patches, a batch size of one, and 256 patches per epoch. The training objective combined Dice loss and binary cross-entropy, balancing region overlap with voxelwise supervision. Optimization used

AdamW with a weight decay of 10^{-4} [12]. The learning rate was reduced from 10^{-3} to 10^{-5} using a cosine schedule over 200 epochs [11]. A 30-subject validation cohort was used for development-stage benchmarking, and final evaluation was performed on an independent held-out 28-subject test cohort.

3 Experiments

We used 132 training, 30 validation, and 28 test subjects from the MRI and tumor-mask component of the CoRe-BT dataset [6]. Subjects missing any required sequence were excluded. Inputs were T1n, T1c, T2w, and T2-FLAIR. Reported targets were edema, enhancing tumor, and necrotic/non-enhancing tumor. Volumes were cropped, clipped to the 0.5th–99.5th percentile range, and normalized for training.

PID selected T1c+T2-FLAIR as the highest-ranked input pair for this task. We then trained 11 architecturally identical 3D U-Net models (one four-input model, all six pairwise models, and four single-input models). This design tested whether the PID ranking predicted downstream segmentation performance.

The models were evaluated on the held-out set using patient-level macro-averaged Dice, HD95, sensitivity, and precision across tumor regions. Statistical analyses included 10,000 bootstrap resamples, Spearman correlation, paired Wilcoxon tests, Friedman tests with Holm correction, and a -0.03 Dice non-inferiority margin [8]. Shapley values over all 16 input coalitions were computed for the four-input model, replacing omitted inputs with training-set channel means inside the brain mask and zeros outside [17, 15].

The complete experiment used one RTX 3090 with GPU memory capped at 2 GB, four logical CPU cores, and 4 GB of system memory for training and feature extraction. Per-model wall-clock and peak-memory profiling were not part of the evaluation, so measured end-to-end speedup and peak-memory reduction are not reported.

4 Analysis and Results

4.1 PID Selection and Segmentation

PID ranked T1c+T2-FLAIR first with a score of 0.963, followed by T1c+T2w (0.841). Across the six pair models, the PID score was positively associated with test-set Dice ($\rho = 0.600$, $p = 0.208$) and negative HD95 ($\rho = 0.600$, $p = 0.208$). Both associations were directionally consistent with downstream performance, but neither was statistically significant, and the analysis had limited power because only six pairs were available.

On the test cohort, the full model achieved average Dice 0.687 and HD95 16.09. T1c+T2-FLAIR achieved Dice 0.676 and HD95 16.77, making it the strongest pair and second overall (Table 2). The selected pair retained 98.5% of the full-model mean Dice while using half the input channels.

Table 1. Region-specific PID scores for all input-pair combinations. Nec., Ed., and Enh. denote necrotic tumor, edema, and enhancing tumor, respectively.

Input pair	Nec.	Ed.	Enh.	Aggregate	Rank
T1c+T2-FLAIR	**0.789**	**1.106**	**1.032**	**0.963**	**1**
T1c+T2w	0.785	0.777	0.947	0.841	2
T1n+T2-FLAIR	0.745	0.932	0.736	0.793	3
T2w+T2-FLAIR	0.669	0.849	0.651	0.712	4
T1n+T2w	0.662	0.680	0.590	0.642	5
T1n+T1c	0.455	0.409	0.728	0.540	6

Table 2. Test-set performance for the top input configurations. Ed, En, and Ne denote edema, enhancing tumor, and necrotic tumor, respectively. HD95, average Dice, sensitivity, and precision are patient-level macro averages across the three regions.

Inputs	Ch.	HD95	Avg. Dice	Sens.	Prec.	Dice Ed	Dice En	Dice Ne
All four	4	**16.09**	**0.687**	**0.709**	**0.714**	0.550	**0.800**	**0.710**
T1c+T2-FLAIR	2	16.77	0.676	**0.709**	0.699	**0.570**	0.767	0.692
T1c+T2w	2	18.06	0.650	0.670	0.683	0.520	0.721	0.707
T1c	1	21.71	0.589	0.606	0.622	0.534	0.574	0.658
T1n+T1c	2	18.21	0.587	0.594	0.612	0.488	0.613	0.661

The selected-vs.-full Dice difference was -0.010 (95% CI $[-0.052, 0.034]$; paired Wilcoxon $p = 0.0505$). HD95 increased by 0.68 (95% CI $[-4.54, 5.09]$; $p = 0.174$). Sensitivity was nearly identical (difference < 0.001; $p = 0.991$), and precision differed by -0.015 ($p = 0.099$). Dice non-inferiority was not established given our choice of margin.

Model configuration affected Dice ($\chi^2_{10} = 164.63$, $p < 10^{-29}$, Kendall's $W = 0.588$) and HD95 ($\chi^2_{10} = 79.07$, $p < 10^{-12}$, $W = 0.282$). The selected pair significantly outperformed most lower-performing pairwise and single-input models after Holm correction. However, differences between the selected pair and either the full model or T1c+T2w were not significant after correction. Complete results for all 11 configurations are reported in the Appendix.

4.2 Post-hoc Attribution

Shapley analysis ranked T2-FLAIR first (0.336; 95% CI 0.298–0.374) and T1c second (0.258; 95% CI 0.203–0.311). Their interaction was also largest (0.307; 95% CI 0.241–0.373; Fig. 3). Pair interactions differed across the six combinations ($\chi^2_5 = 71.22$, $p < 10^{-13}$, Kendall's $W = 0.509$). This separate model-based analysis agreed with the PID-selected input pair.

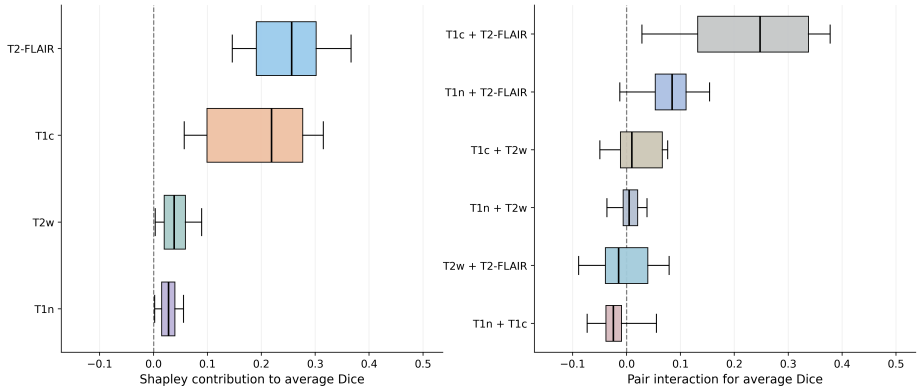

Fig. 3. Test-set Shapley analysis of the four-input model. T2-FLAIR and T1c had the largest input contributions, and their pairwise interaction was the strongest.

5 Discussion

5.1 PID Identified the Strongest Input Pair

The central result is that the MMI-PID score selected the strongest two-input configuration before segmentation training. T1c+T2-FLAIR ranked first by the information score and was subsequently identified as the best pair in the 11-model evaluation. This agreement supports the use of the proposed score as a pre-training input pair selector. The autoencoders and quantizers summarize local image content, while the regional targets connect those summaries to tumor structure before the U-Net is trained.

The selected model used half the input channels and retained 98.5% of the full-input model mean Dice. Paired tests did not establish non-inferiority, and the six available pair configurations provide limited power for a rank-correlation analysis. These results therefore support successful pair selection in this controlled experiment rather than a general statistical claim about ranking performance.

A comparison with simpler training-only criteria is provided in the Appendix. Weighted joint MI and sequential mRMR-MID also selected T1c+T2-FLAIR, whereas ranking inputs only by individual MI selected the substantially weaker T2w+T2-FLAIR pair. This does not establish that the PID weighting universally outperforms simpler MI criteria in a four-input testbed. The methodological contribution of PID is its explicit region-aware decomposition of target-relative information into redundant, unique, and synergistic components, providing a richer characterization of pairwise information structure than a single relevance score.

5.2 The Selected Pair Matches the Imaging Task

The selection of T1c and T2-FLAIR is consistent with the reported regions. T1c directly emphasizes enhancing tumor, while T2-FLAIR is sensitive to edema and surrounding tissue abnormality. Their combination therefore covers two distinct components of the segmentation target. The region-specific PID scores reflected this pattern (Table 1), with T1c+T2-FLAIR achieving the highest aggregate score.

The clinical intuitiveness of the selected pair was also useful as deliberate external face-validity in this proof-of-concept. The objective of the controlled four-contrast experiment was not to discover a clinically surprising pair, but to test whether an automatic training-only information method would recover a compact combination consistent with known imaging characteristics. A stronger test of added value beyond expert selection will require larger and less obvious candidate sets.

Shapley attribution provided a separate model-based audit. T2-FLAIR and T1c had the two largest individual contributions, and their interaction was larger than every other pair interaction. However, PID and Shapley address different questions. PID ranks inputs before downstream training, whereas Shapley measures how a specific trained full model uses its inputs under a defined missing-input baseline. Their agreement strengthens the interpretation that T1c+T2-FLAIR captures both strong individual signal and useful cross-input dependence.

5.3 Implications for Resource-Constrained Training

Selection occurs before costly segmentation training. For n candidate inputs, the method trains n shallow autoencoders and evaluates $\binom{n}{2}$ discrete pair scores, avoiding the need to train a separate downstream segmentation model for every pair during selection. Exhaustive pair training was performed here only to validate the selector.

Reducing four input channels to two also halves the raw input tensor size and the input-dependent parameters and operations of the first convolution. It does not imply a 50% reduction in total U-Net memory, runtime, or computation because later feature maps are unchanged. End-to-end GPU memory, runtime, and total computational savings were not directly profiled in this study. We therefore restrict the resource claim to these direct input-level reductions and to the avoided search cost rather than reporting an unmeasured end-to-end speedup.

The current study remains limited to one dataset, one train/validation/test split, one downstream architecture, four candidate MRI contrasts, and a 28-subject test cohort. The selector was evaluated with a 32-dimensional latent representation, $K = 16$ K-means quantization, and 70% foreground-centered patch sampling. For the evaluated PID terms, T1c+T2-FLAIR remains rank 1 for every $\lambda \geq 0$, including $\lambda = 0$. Sensitivity to alternative K, latent dimensions, sampling fractions, and random seeds was not evaluated in this study. Future work should therefore evaluate these representation choices, additional PID estimators, repeated splits, larger and multisite cohorts, more candidate inputs, and multiple downstream architectures.

6 Conclusion

In this work, a framework using shallow autoencoder encoding, latent quantization, and a region-aware MMI-PID score identified the strongest two-input configuration before segmentation training. The framework selected T1c+T2-FLAIR, which halved the number of input channels, retained 98.5% of the full-model test Dice, and was supported by a separate Shapley analysis. This study provides a focused proof-of-concept for PID-based input selection as a practical and interpretable pre-training strategy for reducing multi-contrast 3D input complexity before costly model development. Broader validation is required to establish robustness across datasets, architectures, representation choices, and larger candidate sets.

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 [14], 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 3. 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 λ sensitivity can be evaluated exactly from the PID terms without re-estimating representations. Equation 2 is affine in λ. 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 4 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 4. 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

References

1. Barrett, A.B.: Exploration of synergistic and redundant information sharing in static and dynamical Gaussian systems. Physical Review E 91(5), 052802 (2015). https://doi.org/10.1103/PhysRevE.91.052802
2. Bertschinger, N., Rauh, J., Olbrich, E., Jost, J., Ay, N.: Quantifying unique information. Entropy 16(4), 2161–2183 (2014). https://doi.org/10.3390/e16042161
3. Chopra, A.S., Neher, C., Ren, T., Heras Rivera, J.E., Jahanian, H., Kurt, M.: SFL-Net: Source-factorized latent representation learning for multi-contrast MRI to tau-PET synthesis. arXiv preprint arXiv:2602.22545 (2026). https://doi.org/10.48550/arXiv.2602.22545, https://arxiv.org/abs/2602.22545
4. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. Lecture Notes in Computer Science, vol. 9901, pp. 424–432. Springer (2016). https://doi.org/10.1007/978-3-319-46723-8_49
5. Cover, T.M., Thomas, J.A.: Elements of Information Theory. John Wiley & Sons, Hoboken, NJ, 2nd edn. (2006)

6. Heras Rivera, J.E., Low, D.K., Xiong, X., Ruzevick, J.J., Child, D.D., Yim, W.w., Kurt, M., Ben Abacha, A.: CoRe-BT: A multimodal radiology-pathology-text benchmark for robust brain tumor typing. arXiv preprint arXiv:2603.03618 (2026). https://doi.org/10.48550/arXiv.2603.03618, https://arxiv.org/abs/2603.03618
7. Hinton, G.E., Salakhutdinov, R.R.: Reducing the dimensionality of data with neural networks. Science 313(5786), 504–507 (2006). https://doi.org/10.1126/science.1127647
8. Holm, S.: A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics 6(2), 65–70 (1979)
9. 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, 203–211 (2021). https://doi.org/10.1038/s41592-020-01008-z
10. Lloyd, S.: Least squares quantization in PCM. IEEE Transactions on Information Theory 28(2), 129–137 (1982). https://doi.org/10.1109/TIT.1982.1056489
11. Loshchilov, I., Hutter, F.: SGDR: Stochastic gradient descent with warm restarts. In: International Conference on Learning Representations (2017)
12. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: International Conference on Learning Representations (2019)
13. Menze, B.H., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Transactions on Medical Imaging 34(10), 1993–2024 (2015). https://doi.org/10.1109/TMI.2014.2377694
14. Peng, H., Long, F., Ding, C.: Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(8), 1226–1238 (2005). https://doi.org/10.1109/TPAMI.2005.159
15. Ren, T., Heras Rivera, J.E., Oswal, H., Pan, Y., Chopra, A.S., Ruzevick, J.J., Kurt, M.: Here comes the explanation: A Shapley perspective on multi-contrast medical image segmentation. In: Joint Proceedings of the xAI 2025 Late-Breaking Work, Demos and Doctoral Consortium. CEUR Workshop Proceedings, vol. 4017, pp. 185–192 (2025), https://ceur-ws.org/Vol-4017/paper_24.pdf
16. Sadegheih, Y., Merhof, D.: Segmentation of brain metastases in MRI: A two-stage deep learning approach with modality impact study. In: Rekik, I., Adeli, E., Park, S.H., Cintas, C. (eds.) Predictive Intelligence in Medicine. Lecture Notes in Computer Science, vol. 15155, pp. 196–206. Springer (2024). https://doi.org/10.1007/978-3-031-74561-4_17
17. Shapley, L.S.: A value for n-person games. In: Kuhn, H.W., Tucker, A.W. (eds.) Contributions to the Theory of Games II, pp. 307–317. No. 28 in Annals of Mathematics Studies, Princeton University Press, Princeton, NJ (1953)
18. Westphal, C., Hailes, S., Musolesi, M.: Partial information decomposition for data interpretability and feature selection. In: Li, Y., Mandt, S., Agrawal, S., Khan, E. (eds.) Proceedings of the 28th International Conference on Artificial Intelligence and Statistics. Proceedings of Machine Learning Research, vol. 258, pp. 1873–1881. PMLR (2025), https://proceedings.mlr.press/v258/westphal25a.html
19. Williams, P.L., Beer, R.D.: Nonnegative decomposition of multivariate information. arXiv preprint arXiv:1004.2515 (2010). https://doi.org/10.48550/arXiv.1004.2515, https://arxiv.org/abs/1004.2515
20. Wollstadt, P., Schmitt, S., Wibral, M.: A rigorous information-theoretic definition of redundancy and relevancy in feature selection based on (partial) information

decomposition. Journal of Machine Learning Research 24(131), 1–44 (2023), https://www.jmlr.org/papers/v24/21-0482.html

21. Zhou, T., Canu, S., Vera, P., Ruan, S.: Feature-enhanced generation and multi-modality fusion based deep neural network for brain tumor segmentation with missing MR modalities. Neurocomputing 466, 102–112 (2021). https://doi.org/10.1016/j.neucom.2021.09.032