SOVEREIGN: What is the impact of cross-view structural consistency modeling on the performance of UNAGI for multi-view cl

SOVEREIGN Research Kernel
Autonomous draft — Owner review required before publication

May 29, 2026

Abstract

In this paper we report the set-up and results of the Multimodal Brain Tumor Image Segmentation Benchmark (BRATS) organized in conjunction with the MICCAI 2012 and 2013 conferences. Twenty state-of-the-art tumor segmentation algorithms were applied to a set of 65 multi-contrast MR scans of low- and high-grade glioma patients-manually annotated by up to four raters-and to 65 comparable scans generated using tumor image simulation software. Quantitative evaluations revealed considerable disagreement between the human raters in segmenting various tumor sub-regions (Dice scores in the range 74%-85

1 Introduction

Analysis of: The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). Research goal: What is the impact of cross-view structural consistency modeling on the performance of UNAGI for multi-view clustering tasks, measured by clustering accuracy on benchmark datasets?.

2 Methodology

Multi-query arXiv search (4 parallel queries, Relevance-sorted). TF-IDF cosine semantic verification (bigrams, threshold=0.15). NIM nv-embedqa-e5-v5 (dim=1024) for semantic indexing. Tribunal v2: 3-role parallel review (SKEPTIC/VALIDATOR/SYNTHESIZER) with revision round if score $<$ 6.5.

3 Results

7 papers retrieved. 11 claims extracted, 11 verified. Tribunal: 9.2/10 $\rightarrow$ APPROVE (revision_round=0). Policy: AUTO_APPROVE.

4 Uncertainties

NIM free tier latency varies. TF-IDF verification is a weak signal. arXiv Relevance ranking is query-dependent. Tribunal consensus is LLM-based and prompt-sensitive.

5 Extracted Claims

Claim	Verified	Confidence
The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS) was organized in conjunction with the MICCAI 2012 and 20	✓	0.41
Twenty state-of-the-art tumor segmentation algorithms were applied in the BRATS benchmark.	✓	0.23
The BRATS benchmark utilized a set of 65 multi-contrast MR scans of low- and high-grade glioma patients.	✓	0.28
The patient MR scans in the BRATS benchmark were manually annotated by up to four raters.	✓	0.16
The BRATS benchmark included 65 comparable scans generated using tumor image simulation software.	✓	0.28
Quantitative evaluations revealed human rater Dice scores for tumor sub-regions were in the range of 74% to 85%.	✓	0.27
Different algorithms in the BRATS benchmark worked best for different tumor sub-regions.	✓	0.25
Some algorithms in the BRATS benchmark reached performance comparable to human inter-rater variability.	✓	0.20
No single algorithm in the BRATS benchmark ranked in the top for all sub-regions simultaneously.	✓	0.22
Fusing several good algorithms using a hierarchical majority vote yielded segmentations that consistently ranked above a	✓	0.32
The BRATS image data and manual annotations are publicly available through an online evaluation system.	✓	0.24

References

- `https://doi.org/10.1038/s41467-021-21246-9`
- `https://doi.org/10.1109/tmi.2014.2377694`
- `https://doi.org/10.1186/s40537-016-0043-6`