ReportX: The BraTS Clinical Report Dataset

Kevin Marchesini*, Omar Carpentiero*, Livia Del Gaudio*,
Francesco Farioli, Rita Cucchiara, Costantino Grana,
Vittorio Cuculo†, and Federico Bolelli†, ✉

University of Modena and Reggio Emilia, Italy
{name.surname}@unimore.it

Abstract. Large-scale benchmarks such as BraTS have driven progress in brain tumor segmentation, but they provide only masks with limited access to the clinical semantics found in radiology reports. We introduce ReportX, a paired resource of 257 clinical reports aligned to BraTS-GLI-2023 subjects, structured into a rich set of qualitative and quantitative attributes. Qualitative fields are curated by clinicians, while quantitative descriptors are automatically derived via atlas-based localization and geometric computations. We compare our annotation schema to existing report-augmented datasets and show that ReportX provides substantially broader coverage of clinically relevant factors. To exploit this supervision, we encode reports using biomedical language models and incorporate their embeddings as auxiliary semantic guidance for 3D tumor segmentation during training. Experimental results demonstrate that the proposed vision-text alignment improves segmentation performance on standard BraTS metrics, with clinically curated reports providing more consistent improvements than automatically generated or less-structured counterparts. We publicly release the dataset[1] and the source code.[2]

Keywords: Brain Tumor · Clinical Reports · Multimodal Integration

1 Introduction

Multimodal medical imaging datasets are increasingly being released to support the training of deep learning algorithms for clinically relevant tasks. In neuro-oncology, BraTS [4,23] is the most extensive benchmark for brain tumor segmentation [3,27], providing four multiparametric MRI (mpMRI) modalities [18], i.e., FLuid-Attenuated Inversion Recovery (FLAIR), T1-weighted (T1), T1-weighted with contrast enhancement (T1c), and T2-weighted (T2), with corresponding tumor masks, targeting three tumor subregions: enhancing tumor (ET), necrotic tumor core (NCR), and peritumoral edema (ED). However, BraTS, like most of

[1] https://ditto.ing.unimore.it/reportx
[2] https://github.com/AImageLab-zip/ReportX
* Equal contribution. Authors are allowed to list their names first on their CVs.
† Equal supervision.
✉ Corresponding author: federico.bolelli@unimore.it

the 3D segmentation datasets [2,10], is limited to voxel-wise annotations, while in clinical practice, a complete understanding of the pathology is obtained through radiology reports that describe not only the tumor, but also other related findings, such as enhancement patterns, ventricular/ependymal involvement, midline shift, and other contextual abnormalities. Enriching BraTS with structured reports is therefore essential to enable and improve a broader set of clinically grounded tasks, such as text-conditioned learning [7, 26, 36, 38], conditioned image synthesis [19], and report generation [24,35,37]. However, 3D vision-language learning remains underdeveloped, with only a limited number of volume-report datasets and 3D encoders, predominantly in CT datasets [9, 11, 14]. Moreover, existing 3D resources are often either automatically generated, limited in clinical scope, or restricted to coarse descriptions, resulting in incomplete semantic supervision for volumetric models.

Within the BraTS context, only a few efforts have introduced paired textual descriptions. TextBraTS [30] extends BraTS-GLI-2020 with 365 volume-level reports derived from pseudo-reports generated by GPT-4o and subsequently refined by experts. BTReport [15] provides 1,251 synthetic reports for the entire BraTS-GLI-2023 dataset generated from automatically extracted quantitative features formatted by LLMs. For the same dataset, AutoRG [21] releases reports for 230 cases on tumor and ventricular features.

Motivated by the absence of a fully clinician-curated and semantically comprehensive resource, we introduce ReportX, a structured extension of BraTS-GLI-2023 providing complete radiology reports at two complementary levels for 257 cases: *(i)* qualitative reports written by expert neuroradiologists, describing tumor subregions and clinically relevant contextual findings visible in mpMRI; *(ii)* structured quantitative reports automatically derived from tumor masks and expert-defined atlases, validated by clinicians. We evaluate report quality through coverage against standard neuroradiology ontologies and structured clinical templates, assessing consistency via inter-clinician agreement.

To further evaluate the practical value of the proposed reports, we assess their role as semantic supervision for 3D brain tumor segmentation within the BraTS framework. In the medical domain, related strategies have been predominantly investigated in 2D radiology, where paired image-report corpora are used to learn domain-specific visual representations (e.g., ConVIRT [41], GLoRIA [16], MGCA [36]). More recent approaches incorporate frozen-language encoders or geometry-aware regularization schemes to guide visual feature learning (e.g., M-FLAG [22], R-Super [7]). Despite these advances, their extension to volumetric 3D segmentation remains largely unexplored. Therefore, we propose a 3D U-Net-based framework that leverages expert-validated radiology reports as auxiliary semantic guidance during training, while preserving image-only inference. This formulation enables a controlled assessment of whether clinically structured reports provide meaningful semantic regularization beyond mask-only supervision.

In summary, our contributions are threefold: *(i)* we introduce the ReportX dataset, the first clinician-curated report extension of BraTS-GLI-2023, guided by structured templates, enriched with atlas-based spatial descriptors and seg-

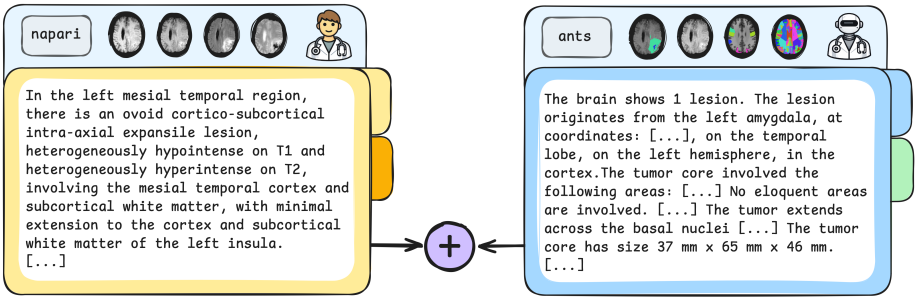


Fig. 1: Overview of the annotation protocol. Clinician reports (left) and automatic reports (right) are independently generated and then concatenated.

mentation-mask-derived geometric measurements; *(ii)* we establish a structured validation protocol assessing report quality in terms of inter-clinician agreement, ontology coverage, and template adherence, demonstrating substantial improvement over existing BraTS report resources; and *(iii)* we design a lightweight 3D vision-text alignment strategy based on frozen biomedical language encoders, to produce report embeddings used only during training, and show that these clinically curated reports provide stronger downstream segmentation gains than automatically generated counterparts.

2 Dataset

Starting from BraTS-GLI-2023 [3], we selected a subset of 257 cases through stratified sampling based on the institution of origin. The selected template followed a standardized internal reporting protocol routinely used in clinical practice by our clinicians and released with the dataset to ensure transparency and reproducibility.

The template fields were divided into two categories: *clinician-annotable* and *automatically-derivable*. This distinction was introduced both to reduce the clinicians' workload and because geometric and quantitative measurements can be computed more accurately and consistently through automated processing. As illustrated in Fig. 1, the two components are subsequently concatenated to form a unified report combining narrative clinical content and quantitative descriptors.

Clinician-Annotable Fields. Two expert clinicians reviewed the cases in Napari, navigating the four imaging modalities. The dataset was divided into two subsets, with an overlap of 50 cases annotated by both clinicians. Reports were written in free-text form while adhering to the predefined structure and required elements of the template to ensure consistency across cases.

Automatically-Derivable Fields. Lesions were identified and counted by computing connected components on the whole tumor mask; cases were labeled as multifocal when distinct tumor core components were present but encompassed within a single contiguous edema region. For anatomical localization, we

Table 1: Agreement between automated fields and clinician annotations.

Field	Prec.	Rec.	F_1	Prev.
Areas (TC)	73.90	96.84	83.83	6.25
Eloq. areas (TC)	80.64	99.02	88.89	0.13
Areas (ED)	84.05	99.70	91.21	12.69

Field	Acc.
Number of lesions	98.00
Origin location	94.23
Origin side	98.36

used multiple atlases in the SRI24 space [29]. A clinician-curated atlas-based on Parc116 [34] was expanded to include fully segmented subcortical and deep white matter regions. Additionally, four binary atlases (vision, speech-motor, speech-receptive, and motor) defined areas where damage may cause significant neurological deficits. These atlases enable automatic identification of the regions where the tumor originated and the anatomical areas involved, including overlap with eloquent regions. All atlases are released together with the dataset. The automatic pipeline relies on the ANTs [33] library and performs deformable image registration using the SyN algorithm to estimate a nonlinear transformation between the atlas T1 and the subject's T1. The resulting subject-specific transformation is applied to the atlases, resulting in a non-rigid deformation that adapts them to the anatomical configuration of each case. Geometric measurements were instead derived from the available segmentation masks. The final 257 complete reports were obtained by appending the automatically derived descriptors to the clinician-written text. As such, the dataset does not belong to the category of fully automatically generated reports, but *reflects a clinician-centered workflow complemented by computational analysis.*

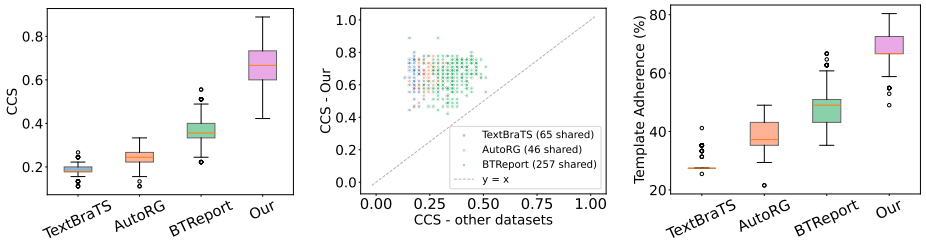
2.1 Dataset Validation

We evaluate the quality of our dataset using three complementary criteria: *(i)* inter-clinician agreement, *(ii)* coverage of standardized brain MRI concepts, and *(iii)* adherence to structured reporting elements.

Inter-Agreement. To quantify the quality of the *clinician-annotated* fields, we analyzed the 50 overlapping reports independently written by both clinicians. Agreement was computed using RadFact [5], which evaluates consistency over structured lists of findings. Each report was decomposed into atomic findings and compared across the two versions. For each finding in report A, the framework checks whether it is supported by a corresponding finding in report B, and the process is repeated in the opposite direction to ensure symmetric evaluation.

Although RadFact is typically applied to compare generated and reference reports, we adapted it to assess agreement between two human-authored reports. RadFact is not limited to chest X-rays; it provides a general methodology for comparing lists of findings. Since both reports are treated as equally valid references, precision and recall are not reported; instead, we report only the symmetric F_1 score. Using LLaMA 3 70B Instruct,[3] we obtained an F_1 of **98.42**.

[3] The Ollama Modelfile is released alongside the dataset.



(a) CCS Box Plot (b) Paired CCS (c) Template Adherence

Fig. 2: Comparison of report quality metrics. (a) CCS distribution across datasets; (b) paired CCS vs. baselines, where points above $y = x$ favor our method; (c) template adherence (%) indicating structural reliability.

For the *automatically-derived* fields, both clinicians annotated the same 50 cases they had in common using the identical structure and predefined queries as the automated pipeline, ensuring a one-to-one correspondence between manual and automatic outputs. This step enabled the evaluation of the accuracy of automatically extracted fields, such as lesion localization and lesion count, against the clinicians' annotations. Table 1 reports precision, recall, F_1, accuracy, and prevalence.

Concept Coverage Score (CCS). To quantify how thoroughly a radiology report describes clinically relevant findings, we introduce the Concept Coverage Score, which measures the proportion of a reference ontology $\mathcal{R}$ addressed within a report. This ontology comprises 45 RadLex concepts derived from the VASARI MRI Feature Set [25] and the RadLex Playbook [28]. Biomedical entities are extracted using GLiNER-BioMed [39] across six domain-specific categories: *anatomical structure*, *lesion*, *tumor characteristic*, *imaging finding*, *MRI sequence*, and *pathological process*. These entities are linked to $\mathcal{R}$ via a dual strategy of exact substring matching and fuzzy string matching (weighted ratio ≥ 85). The CCS is calculated as the cardinality of unique matched concepts normalized by the total size of the reference set. A higher CCS reflects a more comprehensive inclusion of expected neuro-oncological concepts, indicating superior adherence to structured reporting depth.

Template Adherence (TA). To further assess how closely a radiology report follows structured reporting standards for brain tumor MRI, we define the Template Adherence Score. This metric evaluates the presence of 16 clinically relevant features across three hierarchical categories, weighted by their clinical priority: *Clinical Essentials* (6 features, $w = 2.0$) encompassing localization and size [17]; *Morphology and VASARI Features* (7 features, $w = 1.5$) including necrosis and infiltration [13]; and *Advanced/Contextual* features (3 features, $w = 1.0$) such as perfusion and diffusion [8]. Features are identified via regular expression matching. The TA score is calculated as the sum of weights for detected features, normalized by the maximum possible weighted sum across all categories and ex-

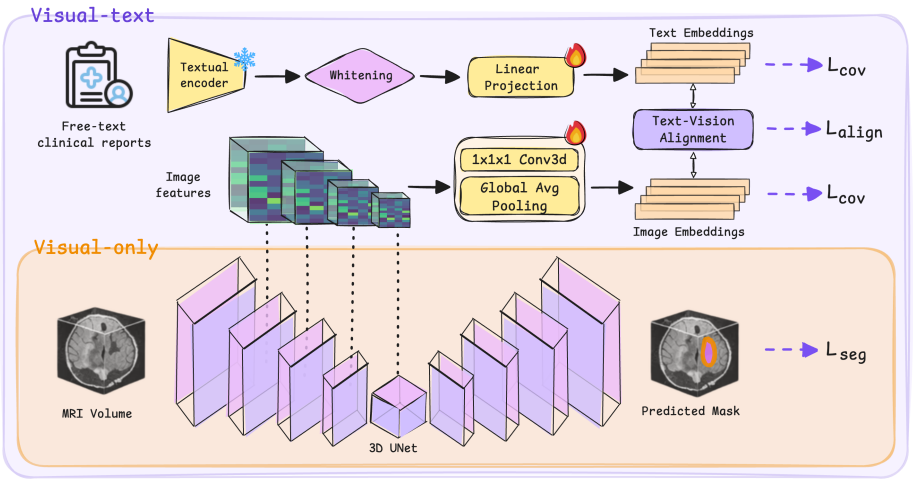


Fig. 3: Our pipeline overview. Encoder features from a 3D U-Net are projected into flat visual embeddings, while clinical reports are mapped to text embeddings. A contrastive vision-text module aligns both modalities during training, while inference relies only on the image backbone.

pressed as a percentage. A higher TA indicates a more exhaustive and organized clinical narrative in alignment with established guidelines.

Validation Results. As shown in Fig. 2, our clinician-annotated reports consistently achieve higher Concept Coverage Scores compared to existing resources, both in distribution (Fig. 2a) and on a per-case basis (Fig. 2b), where most samples lie above the $y = x$ line. Template Adherence scores (Fig. 2c) further confirm stronger compliance with structured reporting standards. Together with the high inter-rater agreement ($F_1 = 98.42$), these results indicate that our dataset provides semantically richer and structurally more consistent reports.

3 Vision-Text Alignment for Segmentation

Method Overview. Beyond intrinsic report quality metrics, we evaluate the practical utility of ReportX in a downstream 3D brain tumor segmentation task, incorporating textual report information as auxiliary supervision during training. The proposed pipeline is reported in Fig. 3. We split the BraTS-GLI-2023 dataset into 875/125/251 cases for train/val/test, and since complete reports are available for 257 training subjects, we alternate two segmentation-only steps with one text-supervised step, where the vision–text alignment branch is activated. This ensures effective contrastive alignment using fully paired image–report batches. At inference time, segmentation relies exclusively on imaging data. We assess segmentation performance through Dice score and 95th percentile Hausdorff Distance (HD95) on the BraTS aggregated regions: Enhancing Tumor (ET), Tumor

Core (TC, comprising ET and NCR), and Whole Tumor (WT, which includes all three tumor subregions).

Joint Vision-Text Representation. Our framework builds upon a 4-level 3D U-Net backbone. Let $f^{(l)} \in \mathbb{R}^{C_l \times H_l \times W_l \times D_l}$ denote the encoder feature tensor at resolution level l, with $l \geq 2$. To construct a compact visual representation, we extract multi-level features and collapse their spatial dimensions via a $1 \times 1 \times 1$ convolution followed by global average pooling. The resulting vectors are aggregated to obtain a visual descriptor $z_v \in \mathbb{R}^d$, with $d = 512$.

To encode clinical reports, we use three frozen BERT-based encoders fine-tuned on medical corpora: BioBERT [20], BioClinical BERT [1], and BioClinical ModernBERT [31]. We observe strong anisotropy in this space, characterized by high pairwise cosine similarity between the extracted text embeddings [12]. To mitigate this effect, embeddings are whitened following [32], and subsequently linearly projected to obtain a compact text embedding $z_t \in \mathbb{R}^d$. The linear projection introduces a trainable transformation in the textual branch, enabling adaptation to the new shared latent space and matching the text embedding dimension to the visual one to compute the contrastive alignment loss.

Vision-Text Alignment Objective. During text-supervised steps, z_v and z_t are fed to the vision-text alignment module enforcing cross-modal consistency through a sigmoid-based contrastive objective. We devised an adapted SigLIP [40] objective: since reports are subject-specific and different subjects may share partially overlapping clinical descriptions, treating non-matching pairs as strict negatives can introduce false negatives; therefore, we consider only positive image–report pairs and rely on VICReg covariance-based regularization [6] to prevent embedding collapse. The training loss is defined as:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{seg}} + \lambda \mathcal{L}_{\text{align}} + \mu \mathcal{L}_{\text{cov}}, \tag{1}$$

where $\mathcal{L}_{\text{seg}}$ denotes the Dice Focal Loss, $\mathcal{L}_{\text{align}}$ the contrastive alignment term, and $\mathcal{L}_{\text{cov}}$ a covariance regularizer applied to the image and text embeddings.

Experimental Settings. Models were trained for 200 epochs, with SGD, on 4 NVIDIA Ampere A100 (64GB) GPUs using PyTorch DistributedDataParallel, with batch size 32 and patch size 128. λ was increased linearly from 0.05 to 0.1 over the first 100 epochs and then linearly decayed to 0, while μ was fixed to 1.

4 Results

Table 2 reports the image-only baseline (no text supervision) and the performance obtained by adding vision-text alignment using our report dataset as well as the three other existing report sources on BraTS [15,21,30]. All numbers are averaged across the three different text encoders, i.e., BioBERT, BioClinical BERT, and BioClinical ModernBERT, to reduce encoder-specific bias. Low standard deviations indicate that our framework is robust to encoder choice. Segmentation conditioned on our reports achieves the best Dice and HD95 scores,

Table 2: Segmentation scores across various report datasets used for vision-text alignment. We report Dice (%) and HD95 (mm) for Base and Whitened embeddings on Enhancing Tumor (ET), Tumor Core (TC) and Whole Tumor (WT).

	Dataset	Dice Score ↑				HD95 ↓			
		ET	TC	WT	Avg.	ET	TC	WT	Avg.
	Baseline	75.8	82.8	93.5	84.0	4.3	7.9	5.2	5.8
Base	TextBraTS	76.5±0.2	83.2±0.1	93.7±0.1	84.5±0.2	4.0±0.0	7.6±0.2	5.5±0.1	5.7±0.0
Base	AutoRG	76.7±0.1	84.0±0.1	**93.9±0.1**	84.8±0.1	4.2±0.2	7.6±0.1	5.3±0.1	5.7±0.1
Base	BTReport	76.7±0.2	**84.4±0.1**	**93.9±0.2**	85.0±0.2	3.9±0.1	7.4±0.0	5.2±0.1	5.5±0.0
Base	**Our ♣**	**78.1±0.0**	**84.4±0.2**	93.7±0.2	**85.4±0.2**	**3.4±0.1**	**6.9±0.1**	**4.9±0.1**	**5.1±0.0**
Whiten.	TextBraTS	76.5±0.4	83.3±0.2	93.5±0.2	84.4±0.2	4.5±0.3	7.6±0.1	5.4±0.3	5.8±0.3
Whiten.	AutoRG	76.9±0.1	84.1±0.0	93.8±0.2	84.9±0.1	4.3±0.3	7.5±0.1	5.1±0.1	5.6±0.2
Whiten.	BTReport	78.4±0.2	84.5±0.0	**94.3±0.0**	85.7±0.1	3.7±0.2	6.9±0.0	4.9±0.1	5.2±0.1
Whiten.	**Our ♣**	**79.1±0.3**	**84.9±0.1**	94.2±0.2	**86.1±0.2**	**3.0±0.2**	**6.0±0.1**	**4.6±0.0**	**4.6±0.2**

Table 3: Ablation study evaluating the impact of alignment loss formulation (standard positive-negative Sigmoid vs positive-only), VIC regularization, and vision-text alignment at different depths on segmentation performance (Dice %).

Loss	Reg	Bottleneck				Multi-level			
		ET	TC	WT	Avg.	ET	TC	WT	Avg.
Standard	✗	76.8±0.1	83.6±0.0	93.6±0.1	84.7±0.0	78.0±0.2	84.1±0.0	93.5±0.1	85.2±0.1
Positive	✗	76.8±0.2	83.7±0.1	93.8±0.1	84.8±0.1	78.3±0.2	84.2±0.1	93.5±0.1	85.3±0.1
Positive	✓	77.7±0.3	84.2±0.0	94.1±0.2	85.3±0.2	**79.1±0.3**	**84.9±0.1**	**94.2±0.2**	**86.1±0.2**

outperforming all other report datasets. These results provide downstream validation of our dataset quality: despite not being the largest source, ReportX yields the largest gains, showing that supervision quality matters more than dataset size. The best overall configuration applies whitening to the text embeddings, improving Dice by over 2 points and reducing HD95 by 1.2, suggesting that mitigating embedding anisotropy enhances cross-modal alignment.

Statistical analysis of the DSC values provides further evidence of the superiority of our dataset. Because the normality assumption was rejected, we applied a non-parametric Friedman test, which was significant ($p = .002$, $\alpha = .05$). The Nemenyi post-hoc test ranked ReportX first in terms of mean rank (1.57), followed by BTReport (2.77), AutoRG (2.96), and TextBraTS (3.35).

Ablations. Table 3 analyzes the contribution of alignment loss, VICReg regularization, and vision-text alignment depth. Extracting visual embeddings from multiple encoder levels consistently outperforms bottleneck-only extraction, confirming that semantic supervision benefits from multi-scale interaction with hierarchical features. When paired with VICReg, the positive-only alignment objective is more stable and achieves consistently higher Dice than the symmetric

formulation, suggesting that explicit negative repulsion is unnecessary for discrimination in our setting. Overall, the best configuration combines multi-level features, positive alignment loss, and VICReg.

5 Conclusion

We introduced a clinically curated extension of BraTS, providing structured, region-aware, and expert-validated textual annotations of brain tumors. Compared to existing report resources, the dataset offers greater clinical consistency and completeness. When used as auxiliary supervision for segmentation, the proposed higher-quality reports translate into consistent improvements in Dice and HD95, outperforming available competitors. This demonstrates that improved dataset quality directly translates into stronger downstream performance under semantic supervision. As future work, we plan to extend the annotation effort to the full BraTS-GLI-2023 dataset, enabling broader investigations beyond segmentation, including report generation and cross-modal representation learning.

Acknowledgments. This project has received funding from Fondazione di Modena, through the FAR 2024 (E93C24002080007), from MUR, under the NRRP "Fit4MedRob - Fit for Medical Robotics" (PNC0000007), and from IRCCS Istituto delle Scienze Neurologiche di Bologna (IRCCS-ISNB).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Alsentzer, E., et al.: Publicly Available Clinical BERT Embeddings. In: Clinical Natural Language Processing Workshop (2019)
2. Antonelli, M., et al.: The Medical Segmentation Decathlon. Nature Communications **13**(1) (2022)
3. Baid, U., et al.: Brain Tumor Segmentation, and Cross-Modality Domain Adaptation for Medical Image Segmentation. Lecture Notes in Computer Science, vol. 14669 (2023)
4. Bakas, S., et al.: Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. Scientific Data **4**(1) (2017)
5. Bannur, S., et al.: MAIRA-2: Grounded Radiology Report Generation. arXiv:2406.04449 (2024)
6. Bardes, A., et al.: VICReg: Variance-Invariance-Covariance Regularization for Self-Supervised Learning. arXiv:2105.04906 (2021)
7. Bassi, P.R., et al.: Learning Segmentation from Radiology Reports. In: Medical Image Computing and Computer Assisted Intervention (2025)
8. Bink, A., et al.: Structured Reporting in Neuroradiology: Intracranial Tumors. Frontiers in Neurology **9** (2018)
9. Blankemeier, L., et al.: Merlin: a computed tomography vision–language foundation model and dataset. Nature (2026)
10. Bolelli, F., et al.: Multi-structure segmentation in CBCT volumes: The ToothFairy2 challenge. Medical Image Analysis (2026)
11. Claessens, C., et al.: Scaling Self-Supervised and Cross-Modal Pretraining for Volumetric CT Transformers. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2026)
12. Ethayarajh, K.: How Contextual are Contextualized Word Representations? Comparing the Geometry of BERT, ELMo, and GPT-2 Embeddings. In: Conference on Empirical Methods in Natural Language Processing and the International Joint Conference on Natural Language Processing (2019)
13. Goodkin, O., et al.: Structured reporting of gliomas based on VASARI criteria to improve report content and consistency. BMC Medical Imaging **25**(1) (2025)
14. Hamamci, I.E., et al.: A foundation model utilizing chest CT volumes and radiology reports for supervised-level zero-shot detection of abnormalities. arXiv:2403.17834 **5** (2024)
15. Heras Rivera, J.E., et al.: BTReport: A Framework for Brain Tumor Radiology Report Generation with Clinically Relevant Features. In: International Conference on Medical Imaging with Deep Learning (2026)
16. Huang, S.C., et al.: GLoRIA: A Multimodal Global-Local Representation Learning Framework for Label-efficient Medical Image Recognition. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (2021)
17. Huckhagel, T., Riedel, C.: MRI reporting of gliomas: What neuro-oncology clinicians expect from radiologists. Radiologie **62**(8) (2022)
18. Iv, M., et al.: Current Clinical State of Advanced Magnetic Resonance Imaging for Brain Tumor Diagnosis and Follow Up. In: Seminars in Roentgenology. vol. 53 (2018)
19. Kim, K., et al.: Controllable Text-to-Image Synthesis for Multi-Modality MR Images. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (2024)

20. Lee, J., et al.: BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics **36**(4) (2020)
21. Lei, J., et al.: Interpretable Brain MRI Report Generation Anchored by Lesion Topography. IEEE Journal of Biomedical and Health Informatics (2025)
22. Liu, C., et al.: M-FLAG: Medical Vision-Language Pre-training with Frozen Language Models and Latent Space Geometry Optimization. In: Medical Image Computing and Computer Assisted Intervention (2023)
23. Menze, B.H., et al.: The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). IEEE Transactions on Medical Imaging **34**(10) (2014)
24. Messina, P., et al.: A Survey on Deep Learning and Explainability for Automatic Report Generation from Medical Images. ACM Computing Surveys **54**(10s) (2022)
25. Negro, A., et al.: VASARI 2.0: a new updated MRI VASARI lexicon to predict grading and IDH status in brain glioma. Frontiers in Oncology **14** (2024)
26. Ni, X., et al.: MG-3D: Multi-Grained Knowledge-Enhanced 3D Medical Vision-Language Pre-training. Medical Image Analysis (2024)
27. Pipoli, V., et al.: IM-Fuse: A Mamba-Based Fusion Block for Brain Tumor Segmentation with Incomplete Modalities. In: Medical Image Computing and Computer Assisted Intervention (2025)
28. Radiological Society of North America: RadLex Playbook. `https://playbook.radlex.org/playbook/SearchRadlexAction` (2026), accessed: 2026-02-26
29. Rohlfing, T., et al.: The SRI24 multichannel atlas of normal adult human brain structure. Human Brain Mapping **31**(5) (2010)
30. Shi, X., et al.: TextBraTS: Text-Guided Volumetric Brain Tumor Segmentation with Innovative Dataset Development and Fusion Module Exploration. In: Medical Image Computing and Computer Assisted Intervention (2025)
31. Sounack, T., et al.: BioClinical ModernBERT: A State-of-the-Art Long-Context Encoder for Biomedical and Clinical NLP. arXiv:2506.10896 (2025)
32. Su, J., et al.: Whitening Sentence Representations for Better Semantics and Faster Retrieval. arXiv:2103.15316 (2021)
33. Tustison, N.J., et al.: The ANTsX ecosystem for quantitative biological and medical imaging. Scientific Reports **11**(1) (2021)
34. Tzourio-Mazoyer, N., et al.: Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain. Neuroimage **15**(1) (2002)
35. Valerio, A.G., et al.: From segmentation to explanation: Generating textual reports from MRI with LLMs. Computer Methods and Programs in Biomedicine **270** (2025)
36. Wang, F., et al.: Multi-Granularity Cross-modal Alignment for Generalized Medical Visual Representation Learning. Advances in Neural Information Processing Systems **35** (2022)
37. Wang, Z., et al.: METransformer: Radiology Report Generation by Transformer With Multiple Learnable Expert Tokens. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2023)
38. Xie, Y., et al.: Rethinking masked image modelling for medical image representation. Medical Image Analysis **98** (2024)
39. Yazdani, A., et al.: GLiNER-BioMed: a suite of efficient models for open biomedical named entity recognition. Bioinformatics (2026)
40. Zhai, X., et al.: Sigmoid Loss for Language Image Pre-Training. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (2023)
41. Zhang, Y., et al.: Contrastive Learning of Medical Visual Representations from Paired Images and Text. In: Machine Learning for Healthcare Conference (2022)