# BeViL: Brain Tumor Detection Beyond Visible Lesions Using Secondary Effects via Contrastive Learning

**Anonymous authors**
Paper under double-blind review

## Abstract

Accurate brain tumor detection from Magnetic Resonance Imaging (MRI) remains challenging when lesions are subtle, partially visible, or missed due to imaging artifacts, sparse slice acquisition, and reconstruction errors. Most existing methods primarily rely on visible lesion appearance and feature fusion across modalities, which require high-resolution, artifact-free, and multimodal 3D MRI acquisitions, which are often time-consuming and costly. To address this limitation, we introduce a novel problem formulation for brain tumor detection that relies solely on the tumor-induced secondary effects present in neighboring slices. Using paired FLAIR, T1-weighted, Post-contrast T1 and T2-weighted MRI, modality-specific DenseNet201 classifiers are trained with a joint objective that combines cross-entropy loss and supervised contrastive loss. The contrastive objective aligns class-consistent representations across modalities, enabling effective representation learning by encouraging shared discriminative features that capture indirect tumor-related cues beyond explicit lesion visibility. During inference, each modality is employed independently. Experimental results demonstrate that the proposed framework consistently outperforms single-modality baselines and non-contrastive multimodal training strategies. Using only secondary tumor-induced effects rather than explicit lesion appearance, the proposed approach reduces the risk of misdiagnosis in the case of sparse acquisitions and can be extended to more reliable and early detection of brain tumors.

## 1 Introduction

MRI is central to the diagnosis of brain tumors due to its high soft-tissue contrast and volumetric representation of anatomical structures. Most existing tumor detection and segmentation approaches, particularly those based on deep learning, implicitly assume that the lesion is directly visible within the analyzed slice. This assumption limits the robustness of the model in clinically relevant scenarios where tumor visibility is compromised, such as sparse slice acquisitions, motion or susceptibility artifacts, or partial volume effects. However, neoplastic processes produce structural and intensity-related secondary alterations which extend beyond the tumor core. Edema, mass effect, ventricular deformation, Blood Brain Barrier (BBB) breaking, and subtle shifts in tissue morphology propagate across neighboring slices, embedding indirect but informative signatures of pathology within the three-dimensional context of volume Solar et al. (2022). Often, these secondary effects play an important role in diagnosing metastases and primary tumor lesions Baris et al. (2016); Samani et al. (2021). In this work, we propose a deep learning framework that infers the tumor presence Beyond Visible Lesion (BeViL) using secondary effects observed in neighboring slices, without requiring explicit tumor visibility in the target slice. The key contributions of the work are:

1. **Problem formulation:** The work frames a novel problem of detecting tumors by using only the tumor-induced secondary effects in neighboring slices.
2. **Dataset deriving:** The work derives the multimodal (T1W,Post-contrast T1,T2W,FLAIR) paired dataset required for the formulated problem from a publicly available dataset *BraTS2023* Menze et al. (2014); Bakas et al. (2017b; 2018); Spyridon et al. (2017); Bakas et al. (2017a).



Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

3. **Concept:** Although contrastive learning and tumor detection have been explored extensively in multiple works, only limited number of them view the four modality images as multiple views of same underlying anatomy. Our work introduces new contrastive knowledge-transfer framework for paired multimodal MRI data.
4. **Model architecture:** The work proposes a novel model architecture using the modality specific classifiers, along with the shared projection head and contrastive loss.
5. **Application:** The primary application of the proposed work is to provide meaningful and tumor aware priors to the models used for reconstruction of full 3D MRI volumes from sparsely sampled 2D slices.

## 2 Related Work

Brain tumor detection from MRI has been extensively studied in the context of segmentation, grading, and classification. While segmentation frameworks aim to delineate tumor subregions, classification models focus on determining the presence or absence of pathology at slice or volume level. In clinical practice, however, reliable tumor detection remains challenging when lesions are subtle, partially visible, or missed due to sparse slice acquisition, inter-slice gaps, motion artifacts, or reconstruction errors.

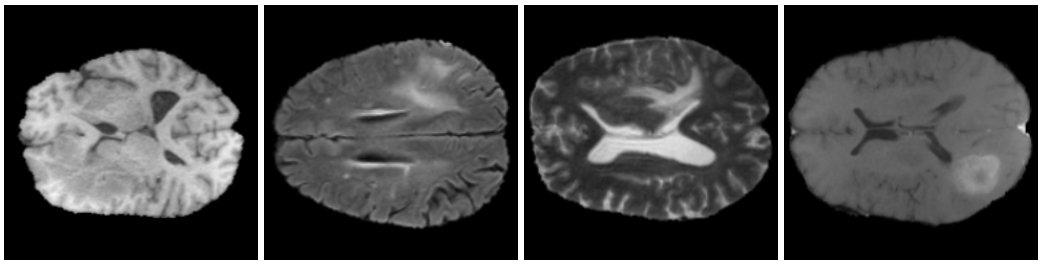
Deep convolutional neural networks (CNNs) have demonstrated strong performance in tumor versus non-tumor classification tasks using both single- and multimodal MRI Rajasree et al. (2021); Tang (2026). Architectures such as VGG, ResNet, DenseNet, EfficientNet, and related variants learn hierarchical representations that capture intensity heterogeneity, contextual asymmetry, and surrounding tissue alterations. Capsule networks have further emphasized spatial relationships and part–whole modeling for tumor classification Paul et al. (2026). More recently, transformer-based architectures and large-scale pre-trained medical vision models have improved global contextual modeling and cross-modal representation learning Abda et al. (2026); Shalam & Korman (2026). Despite these advances, most classification frameworks implicitly assume that tumor cores are sufficiently visible within the analyzed slice or volume. Performance typically degrades when lesions are small, partially captured, or corrupted by acquisition artifacts.

Beyond explicit lesion appearance, tumors induce secondary biological and structural effects, including vasogenic edema, blood–brain barrier disruption, infiltration, tissue compression, and mass effect Alshaikh & Brady (2026); Riascos (2026); Winkler et al. (2023). These indirect manifestations may appear in neighboring slices and across modalities such as FLAIR and T2-weighted MRI, even when the primary tumor core is not fully visible. Large-scale benchmarks such as BRATS provide detailed annotation of tumor subregions and edema Mehta et al. (2022), enabling substantial progress in segmentation research. However, these frameworks primarily optimize delineation accuracy and generally require visible tumor initialization, limiting their applicability in sparse or artifact-prone settings. A related and increasingly important challenge arises in accelerated and sparse-slice MRI acquisition. In clinical workflows, large inter-slice gaps may result in incomplete volumetric coverage, increasing the risk that small tumors fall between acquired slices. Recent diffusion-based generative models have addressed the problem of reconstructing missing information and synthesizing high-fidelity 3D volumes from limited 2D inputs. Score-based diffusion models have been applied to accelerated MRI reconstruction to recover structural details from undersampled data Chung & Ye (2022). Perpendicular 2D diffusion models have been leveraged to enhance volumetric consistency in 3D reconstruction from sparse slices Lee et al. (2023). Cross-sectional diffusion approaches have further demonstrated the feasibility of generating detailed 3D MRI volumes from single or sparse images Bourigault et al. (2024). While these methods improve anatomical continuity and visual realism, their objective is image synthesis rather than pathology-aware detection. Reconstructed volumes may appear structurally plausible, yet reliable tumor classification remains challenging when primary lesion appearance is partially missing or ambiguous.

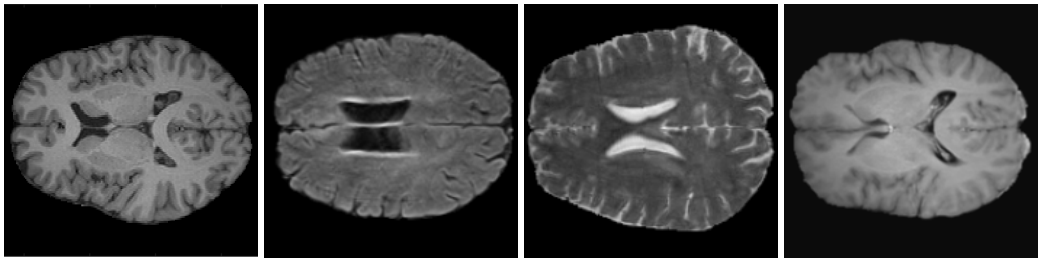
**Contrastive learning** has recently emerged as a powerful strategy for robust representation learning. Supervised contrastive learning promotes intra-class compactness and inter-class separability in embedding space, enhancing generalization under distributional variability Zhang & Wu (2024). Cross-modal contrastive frameworks align paired MRI sequences to learn modality-invariant representations Mok et al. (2024). However, existing studies primarily apply contrastive objectives to segmentation, grading, or domain adaptation tasks, and have not explicitly formulated tumor detec-

tion as a classification problem driven by secondary tumor-induced effects under sparse acquisition conditions.

**Research gap:** Existing tumor classification approaches predominantly rely on explicit lesion visibility and high-quality volumetric imaging, while reconstruction-oriented generative models address structural completeness without directly optimizing pathology detection. A gap therefore remains in developing a classification framework that leverages tumor-induced secondary effects to enable robust detection when lesion cores are missed due to slice gaps. This limitation motivates our supervised multimodal contrastive formulation for slice-level tumor classification under sparse and artifact-prone MRI acquisition settings.



(a) Affected T1 image (b) Affected FLAIR image (c) Affected T2 image (d) Affected T1CE image



(e) Healthy T1 image (f) Healthy FLAIR image (g) Healthy T2 image (h) Healthy T1CE image

Figure 1: Examples of modality-specific tumor-induced secondary effects.

## 3 Dataset Description

We used the BraTS 2023 dataset Menze et al. (2014); Bakas et al. (2017b; 2018); Spyridon et al. (2017); Bakas et al. (2017a), comprising brain MRI volumes from 2,941 patients, including 2,451 patients for model development and 490 independent patients for internal testing. The dataset included four MRI sequences: T1-weighted (T1W), T2-weighted (T2W), FLAIR, and post-contrast T1-weighted (T1CE) imaging. External validation was performed on the independent UTSW-GLIOMA dataset Reddy et al. (2026), comprising MRI scans from all four sequences of 625 tumor patients. Additionally, 20 randomly chosen T1W, 3D MRI scans from Jamaludin et al. (2025) were used for external validation of the model over the healthy subjects.

**Modality-specific secondary effects:** Brain tumors can induce secondary effects such as peritumoral edema and deformation of surrounding anatomical structures, including the ventricular system Ellingson et al. (2015). In Fig 1, the four columns represent the four respective modalities. In the T1-weighted images, Fig. 1a shows deformation of the ventricular structure, whereas Fig. 1e shows the normal ventricular anatomy. In T2-weighted images, increased water content appears hyperintense, making the ventricular boundaries and their deformation more apparent, as shown in Fig. 1c; the corresponding healthy slice is shown in Fig. 1g. FLAIR suppresses the signal from cerebrospinal fluid (CSF), thereby improving the visibility of parenchymal abnormalities and peritumoral edema Ellingson et al. (2015); Jaju et al. (2023). Thus, the hyperintense region in Fig. 1b is consistent with peritumoral edema, whereas Fig. 1f shows no apparent edema. Finally, contrast-enhanced T1-weighted (T1CE) imaging highlights regions of contrast enhancement associated with blood–brain barrier disruption and provides complementary information about tumor morphology Kessler & Bhatt (2018). The image 1d shows the blood-brain-barrier disruption while 1h shows

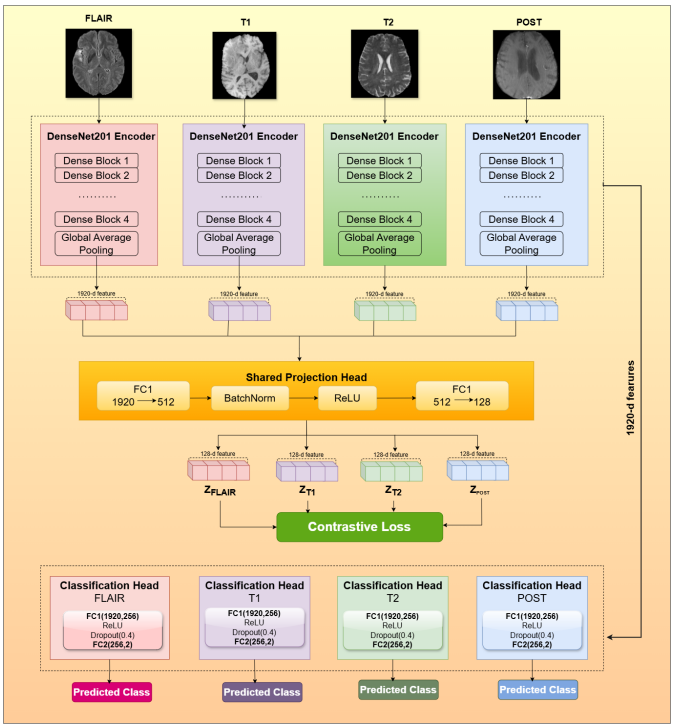



Figure 2: Framework for Modality-specific classification model and contrastive learning.

normal T1CE image. Hence, the four modalities provide complementary information about tumor-associated secondary effects, motivating the transfer of knowledge across modalities.

## 4 METHODOLOGY

This section presents the novel framework for the specific task of detecting the secondary effects of the tumor. The overview of the framework is shown in the Fig 2. The framework is a multi-branch, multimodal deep learning framework designed to exploit the complementary information contained in T1W, T2W, T1CE and FLAIR MRI sequences. The architecture treats each of the modalities as different views of the same underlying anatomy or pathology. Instead of forcing early fusion of heterogeneous signals, it allows each modality to develop its own representation before aligning them in a shared embedding space through contrastive learning. Since each modality contains different biophysical contrast, the framework was designed to develop specialized features.

**Multi-branch feature extractor:** At the top of the architecture are four independent input streams each one corresponding to a modality i.e. T1, T2, FLAIR and T1CE images. Each stream is processed by a dedicated DenseNet201 backbone. In the present framework, each DenseNet201 acts as a modality-specific feature extractor. Because T1 captures anatomical structure and contrast enhancement patterns, T2 emphasizes water content and edema, FLAIR highlights abnormal fluid while suppressing CSF, and T1CE imaging emphasizes contrast-enhancing abnormalities, the four backbones help learn distinct feature hierarchies tailored to these modality-specific signal characteristics.

**Projection Head and Classifier Heads:** Following each DenseNet201 backbones are a modality-specific classification heads. The classification heads consist of two fully connected (FC) layers: FC(1920, 256), followed by a ReLU activation and dropout (0.4), and then FC(256, 2). A shared projection head is employed which contains two fully connected layers: FC(1920, 512), followed by a ReLU activation and dropout (0.4), and then FC(512, 128). It projects the features extracted from modality specific DenseNet201 onto the lower dimensional latent space of size 128.

**Losses:** All the branches of the model are trained simultaneously allowing the gradients of Supervised Contrastive Loss (SupConLoss) to flow through all the three DenseNet201 feature extractors. The overall model is trained using the combined loss function consisting of Binary Cross Entropy Loss (BCE) and SupConLoss. The SupConLoss is defined as follows:

First, we define the pairwise similarity $s_{ij}$ between the features $i$ and $j$

$$s_{ij} = \frac{\mathbf{z}_i^\top \mathbf{z}_j}{\tau} \tag{1}$$

For numeric stability, subtract the row maximum from each $s_{ij}$

$$\tilde{s}_{ij} = s_{ij} - \max_k s_{ik} \tag{2}$$

Considering each feature as an anchor point, define the positives $P(i)$ for that anchor point $i$.

$$P(i) = \{j \mid j \neq i \text{ and } y_j = y_i\} \tag{3}$$

where $y_i$ and $y_j$ are the labels for $i$ and $j$ respectively Finally, the contrastive loss $\ell_i$ for one anchor point $i$ is defined as

$$\ell_i = -\frac{1}{|P(i)|} \sum_{j \in P(i)} \log \left( \frac{\exp(\tilde{s}_{ij})}{\sum_{k \neq i} \exp(\tilde{s}_{ik})} \right) \tag{4}$$

The total contrastive loss for a batch is nothing more than the average of contrastive losses over each anchor point

$$\mathbf{SupConLoss} = \frac{1}{N} \sum_{i=1}^{N} \ell_i, \tag{5}$$

where $N = 3 \times Batch_size$ each batch contains the paired images of T1, T2, and FLAIR modalities.

The BCE is defined as:

$$\mathbf{BCE}_{\text{Mod}} = -\frac{1}{n} \sum_{i=1}^{n} \left[ y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \right], \tag{6}$$

where $n$ is number of samples, $y_i \in \{0, 1\}$ are true labels and $\hat{y}_i \in (0, 1)$ are predicted probabilities.

$$\mathbf{Total_Loss} = \mathbf{BCE_{FLAIR}} + \mathbf{BCE_{T1}} + \mathbf{BCE_{T2}} + \mathbf{BCE_{T1CE}} + \lambda \mathbf{SupConLoss} \tag{7}$$

is the total loss, Where $\lambda$ is the hyperparameter to prevent the inward collapse of the feature space.

### Experimental Setup

The data were divided at the patient level into model-development and internal-test sets, with no patient overlap between the two sets. Each 3D MRI volume had a spatial size of $240 \times 240 \times 155$, corresponding to 155 axial slices of size $240 \times 240$ pixels. Let $T$ denote the set of slice indices that intersect the tumor, with all axial slices indexed from 0 to 154. For a slice indexed by $i$, its distance from the tumor was defined as

$$d(i) = \min_{j \in T} |i - j|. \tag{8}$$

Based on this distance, the primary experiment used a labeling strategy in which slices at distances 1, 2, and 3 from the tumor were labeled positive for tumor presence, while slices at distances 4, 6, and 8 were labeled negative. Note that the images with zero variance in pixels are not added to the dataset. We refer to this labeling strategy as the *three-slice-gap* setting. The model was trained for 50 epochs with $\lambda = 0.3$ and $\tau = 0.7$ using the Adam optimizer with a learning rate of $1 \times 10^{-4}$ and a batch size of 32. The sensitivity analysis of $\lambda$ and $\tau$ is presented in Fig 6 in Appendix A.

## 5 Results

Table 1: Detailed performance analysis of the proposed methodology over primary and external test sets

| Sequence | Metric | Result | 95% CI |
|---|---|---|---|
| **T2-FLAIR** | Accuracy | 0.9156 | (0.9072,0.9242) |
| | Sensitivity | 0.8509 | (0.8357,0.8671) |
| | Specificity | 0.9674 | (0.9600,0.9745) |
| | PPV | 0.9545 | (0.9445,0.9641) |
| | NPV | 0.8900 | (0.8783,0.9018) |
| | Missed-tumor rate | **0.1491** | (0.1329,0.1643) |
| **T2-weighted** | Accuracy | 0.8938 | (0.8839, 0.9029) |
| | Sensitivity | 0.8165 | (0.7978, 0.8343) |
| | Specificity | 0.9558 | (0.9472, 0.9635) |
| | PPV | 0.9368 | (0.9248, 0.9480) |
| | NPV | 0.8665 | (0.8530, 0.8795) |
| | Missed-tumor rate | **0.1835** | (0.1657, 0.2022) |
| **T1-CE** | Accuracy | 0.8819 | (0.8718, 0.8916) |
| | Sensitivity | 0.7892 | (0.7705, 0.8082) |
| | Specificity | 0.9563 | (0.9469, 0.9644) |
| | PPV | 0.9354 | (0.9216, 0.9479) |
| | NPV | 0.8497 | (0.8357, 0.8630) |
| | Missed-tumor rate | **0.2108** | (0.2108, 0.2295) |
| **T1-weighted** | Accuracy | 0.8514 | (0.8401, 0.8621) |
| | Sensitivity | 0.7369 | (0.7165, 0.7564) |
| | Specificity | 0.9433 | (0.9341, 0.9525) |
| | PPV | 0.9125 | (0.8977, 0.9263) |
| | NPV | 0.8172 | (0.8010, 0.8311) |
| | Missed-tumor rate | **0.2631** | (0.2436, 0.2835) |
| **FLAIR(External)** | Accuracy | 0.9433 | (0.9303, 0.9554) |
| | Sensitivity | 0.9467 | (0.9296, 0.9624) |
| | Specificity | 0.9392 | (0.9185, 0.9586) |
| | PPV | 0.9495 | (0.9318, 0.9656) |
| | NPV | 0.9358 | (0.9153, 0.9541) |
| | Missed-tumor rate | **0.0533** | (0.0376, 0.0704) |
| **T2-weighted(External)** | Accuracy | 0.9295 | (0.9141, 0.9417) |
| | Sensitivity | 0.9156 | (0.8930, 0.9352) |
| | Specificity | 0.9463 | (0.9259, 0.9631) |
| | PPV | 0.9537 | (0.9364, 0.9681) |
| | NPV | 0.9027 | (0.8788, 0.9247) |
| | Missed-tumor rate | **0.0844** | (0.0648, 0.1070) |
| **T1-CE(External)** | Accuracy | 0.8630 | (0.8436, 0.8801) |
| | Sensitivity | 0.8548 | (0.8262, 0.8794) |
| | Specificity | 0.8730 | (0.8468, 0.8982) |
| | PPV | 0.8904 | (0.8664, 0.9136) |
| | NPV | 0.8328 | (0.8021, 0.8625) |
| | Missed-tumor rate | **0.1452** | (0.1206, 0.1738) |
| **T1-weighted(External)** | Accuracy | 0.8687 | (0.8493, 0.8866) |
| | Sensitivity | 0.8830 | (0.8596, 0.9073) |
| | Specificity | 0.8515 | (0.8187, 0.8796) |
| | PPV | 0.8778 | (0.8511, 0.9010) |
| | NPV | 0.8577 | (0.8289, 0.8858) |
| | Missed-tumor rate | **0.1170** | (0.0927, 0.1404) |

*Note.* Values are proportions; 95% confidence intervals were calculated using Wilson score intervals. PPV = positive predictive value; NPV = negative predictive value. The missed-tumor rate was calculated as $FN/(TP + FN) = 1 - \text{sensitivity}$.

Table 2: Comparison of Classification Models (Accuracy %)

| Model | FLAIR | T1 | T2 | T1CE |
|---|---|---|---|---|
| Transfusion | 80.28 | 68.14 | 74.07 | 70.42 |
| MedClip | 81.24 | 69.24 | 71.86 | 68.95 |
| ViT-B16 | 81.52 | 63.03 | 74.90 | 72.18 |
| EfficientNetB0 | 82.07 | 71.03 | 72.14 | 73.26 |
| MONAI-DenseNet121 | 83.17 | 68.00 | 76.69 | 74.83 |
| DenseNet121 | 82.34 | 72.97 | 80.96 | 77.64 |
| ConvNeXt-Tiny | 82.62 | 68.55 | 68.69 | 71.37 |
| ResNet50 | 82.76 | 72.14 | 78.48 | 76.92 |
| EfficientNetB3 | 83.17 | 66.76 | 75.45 | 73.84 |
| VGG-19 | 84.00 | 65.10 | 76.97 | 74.51 |
| MobileNetV2 | 84.69 | 69.66 | 78.07 | 76.48 |
| Swin-Transformer | 85.24 | 57.93 | 78.58 | 75.26 |
| DenseNet201 | 84.69 | 72.69 | 79.59 | 78.13 |
| RadDINO | 83.86 | 74.48 | 78.05 | 77.05 |
| **Proposed Model** | **91.56** | **85.14** | **89.38** | **88.19** |

Table 3: Comprehensive ablation study of the proposed framework (Accuracy %).

| Configuration | FLAIR | T1 | T2 | T1CE |
|---|---|---|---|---|
| DenseNet201 (baseline) | 84.69 | – | – | – |
| Pairwise Contrastive (FLAIR–T1) | 85.95 | 71.90 | – | – |
| Pairwise Contrastive (FLAIR–T1–T2) | 87.04 | 75.31 | 83.08 | – |
| **Proposed Model (FLAIR+T1+T2+T1CE)** | **91.56** | **85.14** | **89.38** | **88.19** |

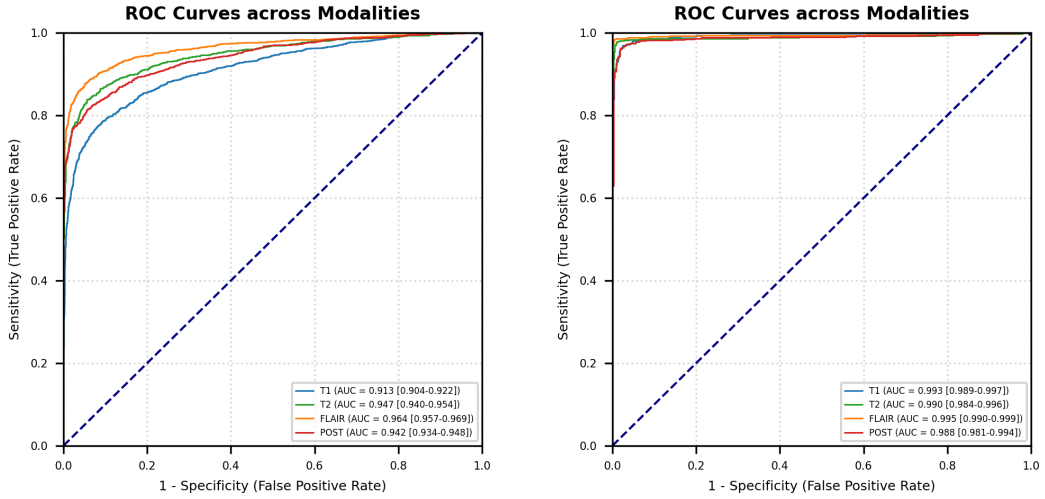
Performance is assessed using accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), missed-tumor rate, and ROC-AUC. The corresponding 95% Wilson confidence intervals are reported to quantify the reliability of the estimates. For external validation, UTSW-GLIOMA dataset Reddy et al. (2026), comprising 625 patients was randomly split into the ratio 10:90(finetuning:testing) at the patient level. The model trained on the primary dataset was finetuned on randomly choosen 10 percent of the external validation dataset for 5 epochs. Table 1 presents the three-slice-gap evaluation on the primary dataset, along with external validation on the independent UTSW-GLIOMA dataset Reddy et al. (2026).

Table 2 presents a quantitative comparison of the proposed model with several classification models across FLAIR, T1W, T2W, and T1CE MRI modalities under the three-slice-gap setting. The proposed model achieves the highest accuracy across all four modalities. Moreover, we also position our method against the baseline fusion methods including *feature fusion* and *late fusion*. The feature fusion and late fusion methods achieves the combined accuracies 87.62% and 87.80% respectively. Note that the proposed method only requires scan from single modality for the inference whereas, the fusion models require scans from all four modalities. More details about the model training and performance of these fusion models are provided in Appendix section A.1. Table 3 presents an **ablation study** examining the contribution of the individual components of the proposed framework. In particular, the final model outperforms models trained using contrastive loss between only two or three modalities, demonstrating the benefit of applying contrastive learning across all four modalities for representation learning.

We further evaluate the model on completely tumor-free MRI data from the Jamaludin et al. (2025) dataset. For this evaluation, performance is measured as the ratio of correct predictions to the total number of predictions. The model achieves an accuracy of 93.24% on T1W images.

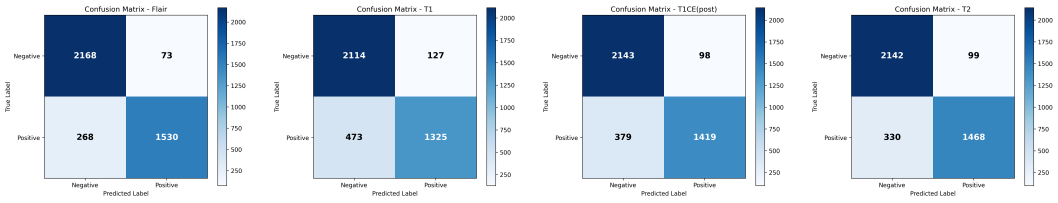
Fig. 3 presents the ROC curves and corresponding AUC values of the proposed model for each modality under the three-slice-gap setting on both the internal and external test sets. The corresponding confusion matrices for the four MRI sequences are shown in Fig. 4. Furthermore, the Grad-CAM visualizations in Fig. 5 show that the model focuses on image regions associated with the secondary effects used for the detection task, providing qualitative evidence that the learned representations capture features relevant to secondary-effect detection.

The effect of the slice-gap labeling strategy is further analyzed in Table 4 in Appendix A. The one- and two-slice-gap settings yield lower performance, likely because their narrow labeling ranges provide limited discriminative information, as tumor-related effects may extend beyond one or two adjacent slices. In contrast, larger gaps improve classification accuracy at the expense of spatial specificity. For example, with a six-slice gap, a positive prediction only indicates the possible presence of a tumor within a 12-slice neighborhood, substantially reducing slice-level localization. We therefore select the three-slice-gap setting as a compromise between class discriminability and spatial specificity, and use it as the primary labeling strategy for subsequent experiments and baseline comparisons.




(a) Internal test set. (b) External test set.

Figure 3: ROC curves and corresponding AUC values of the proposed model on the internal and external test sets.




(a) FLAIR (b) T1 (c) T1CE (d) T2

Figure 4: Confusion matrices for the proposed model with three slice-gaps across FLAIR, T1, T1-CE, and T2 modalities.

## 6 Application and Limitation

Multiple works have shown great results in reconstructing isotropic MRI volumes from sparsely sampled axial slices Pyati et al. (2026). Although the works demonstrates very good quantitative and qualitative results, the risk of small tumors not being captured by the sparsely sampled slices remains. Motivated by this, we introduce a framework which can detect tumors only based on the secondary effects present in the neighboring slices which does not directly capture the tumor. The work can be used to provide useful secondary effects aware priors for such reconstruction models.

The experimental results show that the model works better with three slice-gaps and miserably fails with one slice-gap as shown in Table 4 in Appendix A. This indicates that, for any given slice, the model can only infer whether the tumor is present within three neighboring slices above or below the given slice. The model is unable to determine the exact location of the tumor in terms of direction (above or below) and distance (the number of slices in between) from the given slice. Although the framework uses images from the four modalities independently at inference time, training still requires paired scans from all four modalities. Even though Grad-CAM provides useful information about the features the model focuses on in each image and

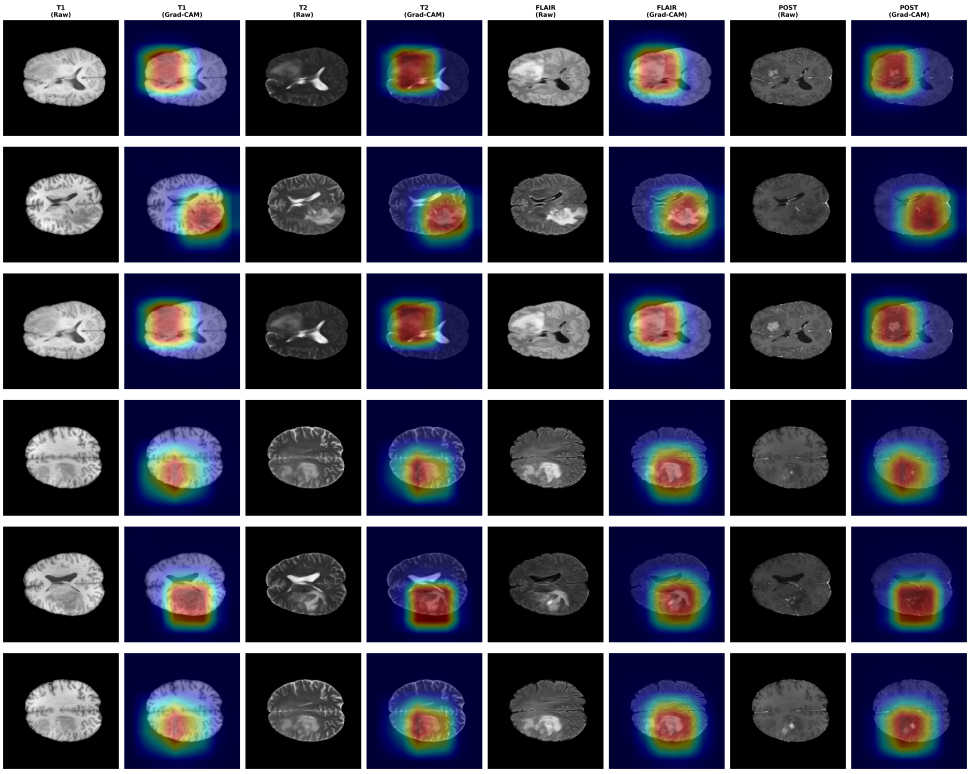



Figure 5: Grad-CAM attention maps across paired FLAIR, T1W, T2W, and T1CE(Post) MRI slices for three slice gap settings.

shows good results in focusing on the relevant region, the distance-based labeling strategy remains prone to noise.

## 7 Conclusion and Future Work

In this work, we introduce a novel framework for a novel problem of detecting tumor presence based only on the secondary effects induced by tumors in neighboring slices. This framework enables reliable brain tumor detection even when the tumor itself is not clearly visible, making it especially valuable in sparse, artifact-corrupted, or low-quality MRI acquisitions. By leveraging secondary tumor-induced effects and contrastive multimodal learning, it supports more robust diagnosis while reducing dependence on costly high-resolution 3D imaging. Moreover, the method introduced in this work can be used to avoid the unnecessary isotropic acquisitions. The experimental results show that by employing the modality specific models with contrastive loss, the model learns the domain specific yet similar representations improving the model performance by transferring knowledge among the modalities.

The future work will focus on developing a framework where the results from this work can be used as priors for the reconstruction model using sparsely sampled slices.

### AI use statement

In this work, there is no use of any kind of generative AI for any kind of tasks. We take responsibility for the final content of this work, including text or claims.

### Ethics statement

Since there is no direct involvement of human or animal subjects in the study, and the dataset used for the study is open access data, there is no ethical concern.

### REPRODUCIBILITY STATEMENT

We assure that rigorous efforts have been made for ensuring the reproducibility of the submitted work. All the tasks have been performed using standard libraries, using fixed seed. Moreover all the training hypermeters are mentioned in the manuscript. All the codes for data derivation, data processing, training, and testing will be made available in the open domain upon the acceptance of the paper to further ensure the reproducibility of the work.

## REFERENCES


Oussama Abda, Hilal Naimi, Mohammed Bourennane, and Shoffan Saifullah. A hybrid wavelet–efficientnetv2–vision transformer framework for multiclass brain tumor mri classification. *Physical and Engineering Sciences in Medicine*, pp. 1–25, 2026.

Jumana Alshaikh and Phil Brady. Brain and neurocognitive effects of radiation. *Aerospace Medicine & Human Performance*, 97(9):679, 2026.

Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, JS Kirby, JB Freymann, Keyvan Farahani, and Christos Davatzikos. Segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection [data set]. the cancer imaging archive. *Version*, 2017a.

Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin S Kirby, John B Freymann, Keyvan Farahani, and Christos Davatzikos. Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific data*, 4(1):170117, 2017b.

Spyridon Bakas, Mauricio Reyes, Andras Jakab, Stefan Bauer, Markus Rempfler, Alessandro Crimi, Russell Takeshi Shinohara, Christoph Berger, Sung Min Ha, Martin Rozycki, et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge. *arXiv preprint arXiv:1811.02629*, 2018.

Mustafa Mahmut Baris, Ahmet Orhan Celik, Naciye Sinem Gezer, and Emel Ada. Role of mass effect, tumor volume and peritumoral edema volume in the differential diagnosis of primary brain tumor and metastasis. *Clinical neurology and neurosurgery*, 148:67–71, 2016.

Emmanuelle Bourigault, Abdullah Hamdi, and Amir Jamaludin. X-diffusion: Generating detailed 3d mri volumes from a single image using cross-sectional diffusion models. 2024.

Hyungjin Chung and Jong Chul Ye. Score-based diffusion models for accelerated mri. *Medical image analysis*, 80:102479, 2022.

Benjamin M Ellingson, Martin Bendszus, Jerrold Boxerman, Daniel Barboriak, Bradley J Erickson, Marion Smits, Sarah J Nelson, Elizabeth Gerstner, Brian Alexander, Gregory Goldmacher, et al. Consensus recommendations for a standardized brain tumor imaging protocol in clinical trials. *Neuro-oncology*, 17(9): 1188–1198, 2015.

Alok Jaju, Yi Li, Hisham Dahmoush, Nicholas G Gottardo, Suzanne Laughlin, David Mirsky, Ashok Panigrahy, Noah D Sabin, Dennis Shaw, Phillip B Storm, et al. Imaging of pediatric brain tumors: a cog diagnostic imaging committee/spr oncology committee/aspnr white paper. *Pediatric blood & cancer*, 70:e30147, 2023.

Iqbal Jamaludin, Mohd Zulfaezal Che Azemin, Mohd Izzuddin Mohd Tamrin, and Abdul Halim Sapuan. "skull-stripped t1-weighted mri of huffaz and non-huffaz brains", 2025.

Alexander T Kessler and Alok A Bhatt. Brain tumour post-treatment imaging and treatment-related complications. *Insights into imaging*, 9(6):1057–1075, 2018.

Suhyeon Lee, Hyungjin Chung, Minyoung Park, Jonghyuk Park, Wi-Sun Ryu, and Jong Chul Ye. Improving 3d imaging with pre-trained perpendicular 2d diffusion models. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 10676–10686. IEEE, 2023.

Raghav Mehta, Angelos Filos, Ujjwal Baid, Chiharu Sako, Richard McKinley, Michael Rebsamen, Katrin Dätwyler, Raphael Meier, Piotr Radojewski, Gowtham Krishnan Murugesan, et al. Qu-brats: Miccai brats 2020 challenge on quantifying uncertainty in brain tumor segmentation-analysis of ranking scores and benchmarking results. *The journal of machine learning for biomedical imaging*, 2022:https–www, 2022.

Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34(10):1993–2024, 2014.

Tony CW Mok, Zi Li, Yunhao Bai, Jianpeng Zhang, Wei Liu, Yan-Jie Zhou, Ke Yan, Dakai Jin, Yu Shi, Xiaoli Yin, et al. Modality-agnostic structural image representation learning for deformable multi-modality medical image registration. In *2024 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp. 11215–11225. IEEE, 2024.

Menali Paul, Ankitesh Tiwari, Bhavesh Barmashe, Kalpana Rai, and Vikas Singh Panwar. Federated mobilenetv2 with ensemble meta-learning for privacy-preserving brain tumor classification. *Scientific Reports*, 2026.

Prajyot Pyati, Sapna Sachan, Amulya Kumar Mahto, and Pranjal Phukan. MK-ResRecon: Multi-kernel residual framework for texture-aware 3d mri refinement from sparse 2d slices. *arXiv preprint arXiv:2605.03432*, 2026.

R Rajasree, C Christopher Columbus, and C Shilaja. Multiscale-based multimodal image classification of brain tumor using deep learning method. *Neural Computing and Applications*, 33(11):5543–5553, 2021.

Divya D. Reddy, N. Saadat, J. M. Holcomb, B. C. Wagner, N. C. Truong, J. Bowerman, K. J. Hatanpaa, T. R. Patel, M. C. Pinho, F. Yu, K. Zhang, S. Lodhi, A. J. Madhuranthakam, C. G. Bangalore Yogananda, and J. A. Maldjian. The university of texas southwestern glioma mri dataset with molecular marker characterization and segmentations (utsw-glioma), 2026.

Roy Riascos. Radiology and artificial intelligence: the radiologist as a leader in the era of medical informatics. *Journal of the Mexican Federation of Radiology and Imaging*, 5(1):1–3, 2026.

Zahra Riahi Samani, Drew Parker, Ronald Wolf, Wes Hodges, Steven Brem, and Ragini Verma. Distinct tumor signatures using deep learning-based characterization of the peritumoral microenvironment in glioblastomas and brain metastases. *Scientific reports*, 11(1):14469, 2021.

Daniel Shalam and Simon Korman. Few-shot biomedical image classification by alignment of independently pretrained encoders. In *ICML 2026 3rd Workshop on Multi-modal Foundation Models and Large Language Models for Life Sciences*, 2026.

Peter Solar, Michal Hendrych, Martin Barak, Hana Valekova, Marketa Hermanova, and Radim Jancalek. Blood-brain barrier alterations and edema formation in different brain mass lesions. *Frontiers in Cellular Neuroscience*, 16:922181, 2022.

Bakas Spyridon, Akbari Hamed, Sotiras Aristeidis, Bilello Michel, Rozycki Martin, Kirby Justin, Freymann John, Farahani Keyvan, and Davatzikos Christos. Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection. *Cancer Imaging Archive*, 2017.

Michael Chi Seng Tang. Enhanced feature extraction technique using multiple pre-trained convolutional neural networks for improved brain tumour detection in mri images. *International Journal of Computational Vision and Robotics*, 17(2):237–246, 2026.

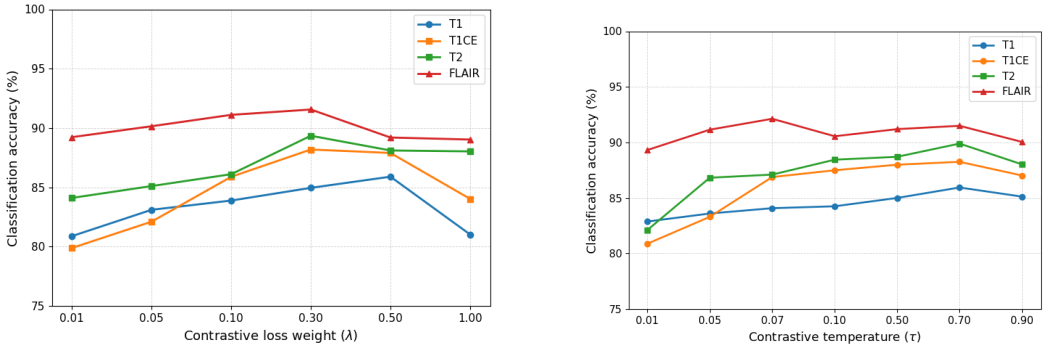
Frank Winkler, Humsa S Venkatesh, Moran Amit, Tracy Batchelor, Ihsan Ekin Demir, Benjamin Deneen, David H Gutmann, Shawn Hervey-Jumper, Thomas Kuner, Donald Mabbott, et al. Cancer neuroscience: State of the field, emerging directions. *Cell*, 186(8):1689–1707, 2023.

Pingyue Zhang and Mengyue Wu. Multi-label supervised contrastive learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 16786–16793, 2024.

# A APPENDIX

Some additional results are given in this section. For this section we define *k slice-gap*( for k = 1, 2, 4, 5, 6) as labelling strategy as follows.

Suppose the slice $s$ has the distance $d$ from the tumor, then $s$ is labeled as positive if $0 < d \leq k$ and $s$ is labeled as negative if $k < d < 3k$ and $d - k$ is odd. The slice $s$ is not included in the dataset if it fails to satisfy both the conditions.




(a) Sensitivity to $\lambda$.

(b) Sensitivity to $\tau$.

Figure 6: Sensitivity of model performance to the hyperparameters $\lambda$ and $\tau$.

Table 4: Performance comparison of different models across 1- to 6-slice gaps in terms of accuracy (%).

| Model | 1-Slice Gap | | | | 2-Slice Gap | | | | 4-Slice Gap | | | | 5-Slice Gap | | | | 6-Slice Gap | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | FLAIR | T1 | T2 | T1CE | FLAIR | T1 | T2 | T1CE | FLAIR | T1 | T2 | T1CE | FLAIR | T1 | T2 | T1CE | FLAIR | T1 | T2 | T1CE |
| Transfusion | 49.96 | 50.18 | 49.87 | 50.12 | 53.62 | 53.41 | 53.62 | 53.28 | 79.84 | 67.52 | 73.21 | 74.18 | 80.11 | 69.03 | 73.84 | 74.63 | 80.56 | 69.47 | 74.12 | 75.02 |
| MedClip | 50.53 | 49.76 | 50.24 | 49.91 | 53.62 | 53.62 | 53.42 | 53.71 | 80.76 | 68.71 | 71.34 | 73.62 | 81.05 | 69.62 | 72.11 | 74.11 | 81.38 | 70.01 | 72.43 | 74.58 |
| ViT-B16 | 50.53 | 50.31 | 49.68 | 50.16 | 53.62 | 53.47 | 53.62 | 53.54 | 81.17 | 62.41 | 74.18 | 75.21 | 81.36 | 63.87 | 74.52 | 75.64 | 81.74 | 64.21 | 75.03 | 76.03 |
| EfficientNetB0 | 49.30 | 50.27 | 49.84 | 50.08 | 51.97 | 53.83 | 52.38 | 52.76 | 81.63 | 70.42 | 71.68 | 73.84 | 81.91 | 71.26 | 72.03 | 74.31 | 82.34 | 71.64 | 72.47 | 74.79 |
| MONAI-DenseNet121 | 50.53 | 49.92 | **50.37** | 50.11 | 53.62 | 53.62 | 53.62 | 53.47 | 82.81 | 67.34 | 75.92 | 76.34 | 83.02 | 68.91 | 76.34 | 76.82 | 83.41 | 69.28 | 76.82 | 77.21 |
| DenseNet121 | 50.29 | 50.14 | 49.73 | 50.28 | 55.90 | 55.07 | 54.45 | 55.21 | 82.01 | 72.21 | 80.34 | 80.91 | 82.48 | 72.86 | 80.71 | 81.34 | 82.79 | 73.42 | 81.13 | 81.72 |
| ConvNeXt-Tiny | 50.53 | 49.81 | 50.06 | 49.94 | 53.62 | 53.62 | 53.62 | 53.62 | 82.31 | 67.83 | 68.21 | 70.84 | 82.57 | 68.34 | 68.84 | 71.37 | 82.91 | 68.76 | 69.13 | 71.81 |
| ResNet50 | 50.53 | 50.22 | 49.91 | 50.34 | 55.90 | 55.07 | 53.62 | 54.18 | 82.43 | 71.48 | 77.91 | 78.32 | 82.68 | 72.51 | 78.24 | 78.76 | 83.04 | 72.93 | 78.76 | 79.18 |
| EfficientNetB3 | 49.88 | 50.16 | 49.79 | 50.21 | 53.42 | 54.66 | 54.87 | 54.31 | 82.86 | 66.12 | 74.91 | 75.62 | 83.14 | 67.01 | 75.28 | 76.13 | 83.46 | 67.43 | 75.67 | 76.52 |
| VGG-19 | 50.53 | 49.85 | 50.13 | 50.07 | 53.62 | 53.62 | 53.83 | 53.62 | 83.71 | 64.42 | 76.48 | 77.24 | 83.94 | 65.37 | 76.82 | 77.68 | 84.26 | 65.84 | 77.21 | 78.05 |
| MobileNetV2 | 49.14 | 50.26 | 49.92 | 50.18 | 54.04 | 53.21 | 54.66 | 54.04 | 84.31 | 68.94 | 77.61 | 78.41 | 84.52 | 69.83 | 78.02 | 78.86 | 84.87 | 70.21 | 78.43 | 79.32 |
| DenseNet201 | 49.96 | 50.11 | 49.82 | 50.23 | 54.04 | 53.42 | 53.83 | 53.83 | 84.28 | 72.04 | 79.17 | 80.12 | 84.51 | 72.84 | 79.83 | 80.71 | 84.82 | 73.26 | 80.14 | 81.03 |
| Swin-Transformer | 49.47 | **50.34** | 49.76 | 50.09 | 53.83 | 53.62 | 53.62 | 53.62 | 84.91 | 57.21 | 77.96 | 78.14 | 85.03 | 58.34 | 78.31 | 78.63 | 85.38 | 58.71 | 78.67 | 79.02 |
| RadDINO | 50.29 | 49.68 | 50.17 | 49.93 | 50.20 | 52.82 | 50.60 | 51.18 | 83.54 | 73.81 | 77.62 | 78.02 | 83.78 | 74.92 | 78.16 | 78.61 | 84.12 | 75.31 | 78.54 | 79.05 |
| **Proposed Model** | **50.53** | 50.27 | 49.91 | **50.34** | **68.64** | **57.59** | **58.38** | **60.84** | **90.66** | **85.49** | **88.24** | **82.47** | **90.65** | **86.04** | **89.47** | **88.86** | **91.01** | **87.88** | **89.52** | **87.23** |

## A.1 FUSION MODELS

Moreover, the four MRI modalities (T1, T2, T2-FLAIR and T1-CE) are independently encoded using DenseNet-201, with multimodal representations integrated through feature-level concatenation (feature fusion) or decision-level logit averaging (late fusion) for binary tumor classification. The models are trained using cross-entropy loss and evaluated on a patient-level test split using accuracy, precision, recall.

Table 5: Performance metrics of Fusion methods methods.

| **Method** | **Accuracy** | **Precision** | **Recall** |
|---|---|---|---|
| Feature fusion | 0.8762 | 0.9295 | 0.8956 |
| Late fusion | 0.8780 | 0.9344 | 0.8930 |