Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Towards UASC-Net: A Pilot Study of Uncertainty-Aware 3D Brain Tumor Segmentation on BraTS 2021

Mojtaba Jahanian

mojtaba160672000@aut.ac.ir

Islamic Azad University of Hamedan

Abbas Karimi

Islamic Azad University, Karaj

Faraneh Zarafshan

Islamic Azad University, Karaj

Alireza Enami

Islamic Azad University, Arak

Maryam hajiee

Islamic Azad University South Tehran Branch

Article

Keywords: Brain tumor segmentation, multimodal MRI, self-supervised learning, deep clustering, uncertainty estimation, semi-supervised learning, pseudo-label refinement

Posted Date: October 5th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10182597/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Towards UASC-Net: A Pilot Study of Uncertainty-Aware 3D Brain Tumor Segmentation on BraTS 2021

Mojtaba Jahanian[a,b,*], Abbas Karimi[c,d,**], Faraneh Zarafshan[c,d], Alireza Enami[e], Maryam hajiee[f]

[a]*Department of Computer Engineering, National University of Skills (NUS), Tehran, Iran*
[b]*Department of Computer, Hmd.C., Islamic Azad University, Hamadan, Iran*
[c]*Department of Computer Engineering, Karaj Branch, Islamic Azad University, Karaj, Iran*
[d]*Institute of Artificial Intelligence and Social and Advanced Technologies, Ka.C., Islamic Azad University, Karaj, Iran*
[e]*Department of Software Engineering, Ar.C., Islamic Azad University, Arak, Iran*
[f]*Department of Computer Engineering, ST.C., Islamic Azad University, Tehran, Iran*

Abstract

Brain tumor segmentation from multimodal magnetic resonance imaging (MRI) is a central component of treatment planning and disease monitoring, but reliable voxel-wise annotation is costly and difficult to obtain. This work presents a pilot evaluation of UASC-Net Lite v2, an uncertainty-aware extension of a compact 3D U-Net baseline for binary tumor segmentation from multimodal BraTS 2021 MRI patches. The model combines a 3D encoder–decoder segmentation pathway with attention-, prototype-, and uncertainty-related auxiliary outputs, enabling spatial confidence inspection alongside mask prediction. Experiments were conducted on a reproducible 16-case pilot subset (13 training and 3 validation cases), using $64 \times 64 \times 64$ patches and eight training epochs. After validation-set threshold calibration, the proposed model achieved a pooled Dice score of 0.7381 and IoU of 0.5849, compared with 0.7297 and 0.5744 for the TinyUNet3D baseline. UASC-Net Lite v2 improved sensitivity from 0.7681 to 0.8609, while specificity and precision changed from 0.9685 and 0.6949 to 0.9532 and 0.6460, respectively. Qualitative visualizations show the segmentation outputs together with uncertainty and attention maps. These results should be interpreted as a pilot study rather than a full benchmark: threshold calibration was performed on the validation patches and the cohort is small. Nevertheless, the experiment provides reproducible evidence that uncertainty-aware modeling can improve lesion recall and overlap under limited-data conditions.

Keywords: Brain tumor segmentation, multimodal MRI, self-supervised learning, deep clustering, uncertainty estimation, semi-supervised learning, pseudo-label refinement

Highlights

*Corresponding author.
**Corresponding author.
Email addresses: Mojtaba160672000@aut.ac.ir (Mojtaba Jahanian), Akarimi@iau.ac.ir (Abbas Karimi), Fa.zarafshan@iau.ac.ir (Faraneh Zarafshan), alireza.enami@iau.ac.ir (Alireza Enami), Mhajiee@iau.ac.ir (Maryam hajiee)

- An uncertainty-aware self-supervised deep clustering framework is proposed for annotation-efficient brain tumor MRI segmentation.
- Prototype-based clustering is used to discover tumor-related latent structures and generate pseudo-labels from unlabeled MRI volumes.
- A reliability-aware refinement module suppresses uncertain pseudo-labels using entropy, prototype distance margins, and stochastic prediction variability.
- The framework supports limited-label training and produces segmentation maps, uncertainty maps, and prototype-level interpretability outputs.

1. Introduction

Gliomas are among the most aggressive primary brain tumors and exhibit substantial spatial, radiological, molecular, and treatment-related heterogeneity. Multimodal magnetic resonance imaging (MRI) is routinely used in neuro-oncology because different sequences emphasize complementary tissue characteristics. T1-weighted contrast-enhanced MRI is particularly informative for enhancing tumor components, whereas T2-weighted and fluid-attenuated inversion recovery (FLAIR) images highlight edema and infiltrative non-enhancing abnormalities. Accurate delineation of tumor subregions is therefore essential for treatment planning, surgical and radiotherapy guidance, response assessment, longitudinal monitoring, and quantitative imaging biomarker extraction (Menze et al., 2015; Correia de Verdier et al., 2024).

Deep learning has become the dominant paradigm for medical image segmentation. The U-Net architecture established the encoder-decoder design with skip connections as a strong foundation for biomedical segmentation (Ronneberger et al., 2015), while 3D U-Net extended this principle to volumetric data and enabled dense prediction from sparsely annotated volumes (Çiçek et al., 2016). More recent frameworks such as nnU-Net showed that careful pipeline configuration, including preprocessing, architecture selection, training schedule, and post-processing, can achieve strong and robust performance across diverse biomedical segmentation tasks (Isensee et al., 2021). Transformer-based architectures have further improved the modeling of long-range spatial dependencies in volumetric segmentation, as demonstrated by UNETR and Swin UNETR (Hatamizadeh et al., 2022b,a). Despite these advances, high-performing segmentation systems remain strongly dependent on dense voxel-level annotations, which are expensive to obtain, require domain expertise, and may vary across raters and institutions.

Annotation scarcity is a particularly important limitation in brain tumor MRI segmentation. Public benchmarks such as BraTS have accelerated method development by providing expert-annotated multimodal MRI data and standardized evaluation protocols (Menze et al., 2015; Correia de Verdier et al., 2024). However, creating dense voxel-level labels for new tumor entities, imaging protocols, treatment stages, and clinical institutions remains costly. In practical scenarios, unlabeled scans are often more abundant than expert annotations. This imbalance motivates learning strategies that can exploit unlabeled data while preserving reliability in clinically sensitive segmentation tasks.

Self-supervised learning (SSL) has emerged as an effective strategy for learning transferable representations from unlabeled images. In natural image analysis, contrastive learning and masked image modeling have demonstrated that strong representations can be learned without manual labels (Chen et al., 2020; He et al., 2022). In medical imaging, SSL is especially attractive because unlabeled scans are routinely available, and recent reviews have emphasized the value of contrastive, restorative, masked, and generative pretext tasks for reducing annotation dependency (Shurrab and Duwairi, 2022; Haghighi et al., 2024; Zeng et al., 2024; Liu et al., 2024). Nevertheless, SSL by itself does not guarantee that learned latent representations are organized into tumor-relevant structures. Moreover, SSL-based fine-tuning still typically requires labeled data and does not explicitly control the reliability of pseudo-supervision generated from unlabeled samples.

Deep clustering and prototype learning provide a complementary mechanism for structuring unlabeled representations. Methods such as Deep Embedded Clustering, DeepCluster, and SwAV showed that feature learning can be coupled with cluster assignments or learnable prototypes to discover meaningful visual patterns without direct labels (Xie et al., 2016; Caron et al., 2018, 2020). In brain tumor MRI, such prototypes may capture recurrent appearance patterns associated with edema, enhancing tumor, necrotic or non-enhancing core, normal tissue, and ambiguous boundary regions. However, cluster assignments are inherently uncertain, especially in heterogeneous lesions and low-contrast transition areas. Directly using all cluster assignments as pseudo-labels can introduce confirmation bias and propagate errors during segmentation training.

Uncertainty estimation offers a principled way to evaluate the reliability of predictions and pseudo-labels. Bayesian interpretations of dropout provide a practical approximation for predictive uncertainty in deep networks (Gal and Ghahramani, 2016). In semi-supervised learning, uncertainty-aware pseudo-label selection has been used to reduce the influence of unreliable pseudo-labels (Rizve et al., 2021), and recent medical image segmentation studies have further shown that uncertainty estimation can improve pseudo-label quality and semi-supervised training stability (Lu et al., 2023; Chen et al., 2024). These observations suggest that uncertainty should not be treated as a post-hoc visualization only, but should be integrated into the pseudo-label generation and filtering process.

Motivated by these gaps, this paper proposes **UASC-Net**, an **U**ncertainty-**A**ware **S**elf-supervised deep **C**lustering network for annotation-efficient brain tumor MRI segmentation. UASC-Net integrates self-supervised volumetric representation learning, prototype-based deep clustering, uncertainty-guided pseudo-label refinement, and semi-supervised segmentation decoding within a unified framework. The method is designed for limited-label settings in which a small subset of MRI volumes has expert annotations while the remaining data are unlabeled. Unlike conventional supervised segmentation models, UASC-Net explicitly exploits unlabeled MRI volumes. Unlike standard pseudo-labeling approaches, it estimates voxel-level reliability before using pseudo-labels for training.

The main contributions of this work are summarized as follows:

- We propose an annotation-efficient brain tumor segmentation framework that combines self-supervised volumetric representation learning with prototype-based deep clustering.
- We introduce a prototype-driven pseudo-label generation mechanism that structures latent MRI features into recurrent tumor-related imaging patterns.

- We design an uncertainty-guided pseudo-label refinement module that combines assignment entropy, prototype distance margin, and stochastic prediction variability to suppress unreliable pseudo-labels.
- We formulate a semi-supervised segmentation training strategy that jointly uses limited expert annotations and uncertainty-filtered pseudo-labels from unlabeled MRI volumes.
- We define a comprehensive evaluation protocol including segmentation accuracy, clustering quality, calibration analysis, pseudo-label reliability, limited-label experiments, and ablation studies.

2. Related Work

2.1. Brain Tumor MRI Segmentation

The BraTS benchmark has played a central role in the development and evaluation of automated glioma segmentation algorithms. The original BraTS benchmark provided multimodal MRI data and standardized annotations for tumor subregions, enabling systematic comparison of segmentation methods (Menze et al., 2015). More recent BraTS challenges continue to expand the benchmark toward clinically relevant scenarios, including post-treatment glioma segmentation, and emphasize standardized evaluation using Dice similarity coefficient and Hausdorff distance (Correia de Verdier et al., 2024). These datasets have encouraged the development of robust segmentation pipelines and have made brain tumor segmentation one of the most widely studied tasks in medical image computing.

CNN-based encoder-decoder networks remain strong baselines for this task. U-Net introduced the contracting-expanding architecture with skip connections for precise localization (Ronneberger et al., 2015), and 3D U-Net adapted this design to volumetric biomedical segmentation (Çiçek et al., 2016). nnU-Net further demonstrated that a self-configuring pipeline can outperform many manually designed methods by systematically adapting preprocessing, network configuration, and training details to each dataset (Isensee et al., 2021). Transformer-based models have subsequently improved global context modeling in 3D medical imaging. UNETR introduced a transformer encoder for volumetric segmentation (Hatamizadeh et al., 2022b), while Swin UNETR used hierarchical shifted-window transformers and achieved strong performance in brain tumor MRI segmentation (Hatamizadeh et al., 2022a). Nevertheless, these methods are primarily supervised and generally rely on substantial voxel-level annotations.

2.2. Self-Supervised Learning in Medical Image Analysis

Self-supervised learning aims to learn representations from unlabeled data by defining supervisory signals from the data itself. Contrastive learning methods such as SimCLR learn invariant representations by pulling together augmented views of the same image and pushing apart different images (Chen et al., 2020). Masked image modeling, exemplified by masked autoencoders, learns representations by masking a large fraction of image patches and reconstructing missing content (He et al., 2022). These strategies are appealing in medical imaging because large unlabeled imaging archives are often available, whereas expert annotations are limited.

In medical image analysis, SSL has been investigated through contrastive, restorative, predictive, generative, and masked reconstruction tasks. Recent reviews report that SSL can improve downstream performance and reduce annotation requirements across classification, detection, and segmentation tasks (Shurrab and Duwairi, 2022; Haghighi et al., 2024; Zeng et al., 2024; Liu et al., 2024). For multimodal MRI, SSL is particularly relevant because anatomical structures and pathological regions are spatially aligned across modalities while exhibiting different intensity characteristics. However, many SSL methods focus on representation pretraining and still rely on supervised fine-tuning for segmentation. They do not explicitly convert unlabeled representations into structured pseudo-labels, nor do they evaluate the reliability of generated supervision.

2.3. Deep Clustering and Prototype Learning

Deep clustering jointly learns feature representations and cluster assignments. DEC introduced a framework for learning low-dimensional embeddings and iteratively refining cluster assignments using a clustering objective (Xie et al., 2016). DeepCluster showed that clustering assignments can be used as pseudo-supervision for training visual representations (Caron et al., 2018). SwAV extended this idea by contrasting cluster assignments across different views and using prototypes for efficient self-supervised representation learning (Caron et al., 2020). These methods demonstrate that prototypes can impose structure on latent representations without manual labels.

In medical imaging, prototype learning and clustering are useful because pathological tissue appearances are heterogeneous and may not be captured by a single compact class distribution. For brain tumors, a final semantic class such as edema or tumor core may contain multiple visual subpatterns caused by anatomical variation, scanner differences, treatment effects, and tumor heterogeneity. Prototype-based clustering can therefore provide a more flexible latent partition than direct class prediction. However, a key limitation is cluster-to-class ambiguity. Clusters are not guaranteed to correspond to clinically meaningful tissue types, and unreliable clusters near lesion boundaries may degrade training if used indiscriminately. This motivates the combination of prototype clustering with limited expert supervision and uncertainty-aware pseudo-label refinement.

2.4. Semi-Supervised Segmentation and Pseudo-Label Reliability

Semi-supervised learning aims to use both labeled and unlabeled data. In segmentation, common strategies include consistency regularization, teacher-student learning, co-training, and pseudo-labeling. The Mean Teacher framework showed that consistency between student and teacher predictions can improve semi-supervised learning (Tarvainen and Valpola, 2017). Pseudo-labeling extends this principle by using confident predictions on unlabeled data as additional supervision. However, noisy pseudo-labels can introduce confirmation bias, particularly when the model is overconfident but wrong.

Uncertainty-aware pseudo-label selection has been proposed to reduce this risk. Uncertainty-aware pseudo-labeling methods select or weight pseudo-labels according to prediction confidence (Rizve et al., 2021). In medical image segmentation, uncertainty-aware consistency and pseudo-label refinement have been used to improve the reliability of unlabeled supervision (Lu et al., 2023; Chen et al., 2024). Predictive uncertainty is especially important in tumor

segmentation because errors often occur around irregular boundaries, low-contrast lesion regions, and ambiguous tissue transitions. UASC-Net builds on these insights by estimating uncertainty at the pseudo-label refinement stage rather than using clustering assignments directly.

3. Proposed Method

3.1. Overview of UASC-Net

This study proposes UASC-Net, an uncertainty-aware self-supervised deep clustering framework for annotation-efficient brain tumor MRI segmentation. The framework is designed to exploit both unlabeled and sparsely labeled multimodal MRI data while reducing the propagation of noisy pseudo-labels during training. As shown in Fig. 1, UASC-Net consists of five main components: (1) multimodal MRI input and preprocessing, (2) self-supervised representation learning, (3) prototype-based deep clustering, (4) uncertainty-guided pseudo-label refinement, and (5) semi-supervised segmentation decoding.

The central idea is to first learn transferable volumetric representations from unlabeled multimodal MRI using a self-supervised encoder. These latent representations are then organized through prototype-based clustering to identify recurrent tumor-related imaging patterns and generate initial pseudo-labels. Since cluster-derived pseudo-labels may be unreliable in heterogeneous or low-contrast regions, an uncertainty-guided refinement module estimates voxel-level reliability and suppresses uncertain assignments. Finally, the refined pseudo-labels are combined with limited expert annotations to train a segmentation decoder in a semi-supervised manner. This design allows the model to benefit from structural information available in unlabeled MRI volumes while maintaining robustness against pseudo-label noise.

Table 1 summarizes the main mathematical notation used in the proposed method.

3.2. Multimodal MRI Input and Preprocessing

Each subject is represented by a multimodal MRI volume

$$X \in \mathbb{R}^{C\times D\times H\times W}, \tag{1}$$

where $C = 4$ corresponds to T1, T1ce, T2, and FLAIR MRI modalities, and D, H, and W denote the depth, height, and width of the 3D volume. When voxel-level annotation is available, the segmentation mask is denoted as

$$Y \in \{0, 1, \ldots, K-1\}^{D\times H\times W}, \tag{2}$$

where K is the number of segmentation classes. In BraTS-style evaluation, voxel-level classes are also aggregated into clinically meaningful regions such as whole tumor, tumor core, and enhancing tumor (Menze et al., 2015; Correia de Verdier et al., 2024).

Representative examples of the multimodal input data are shown in Fig. 2. The figure illustrates the complementary appearance of tumor-related regions across MRI modalities. In particular, T1ce commonly provides strong contrast for enhancing tumor regions, while T2 and FLAIR are more informative for edema and non-enhancing abnormalities. This complementarity motivates joint multimodal representation learning.

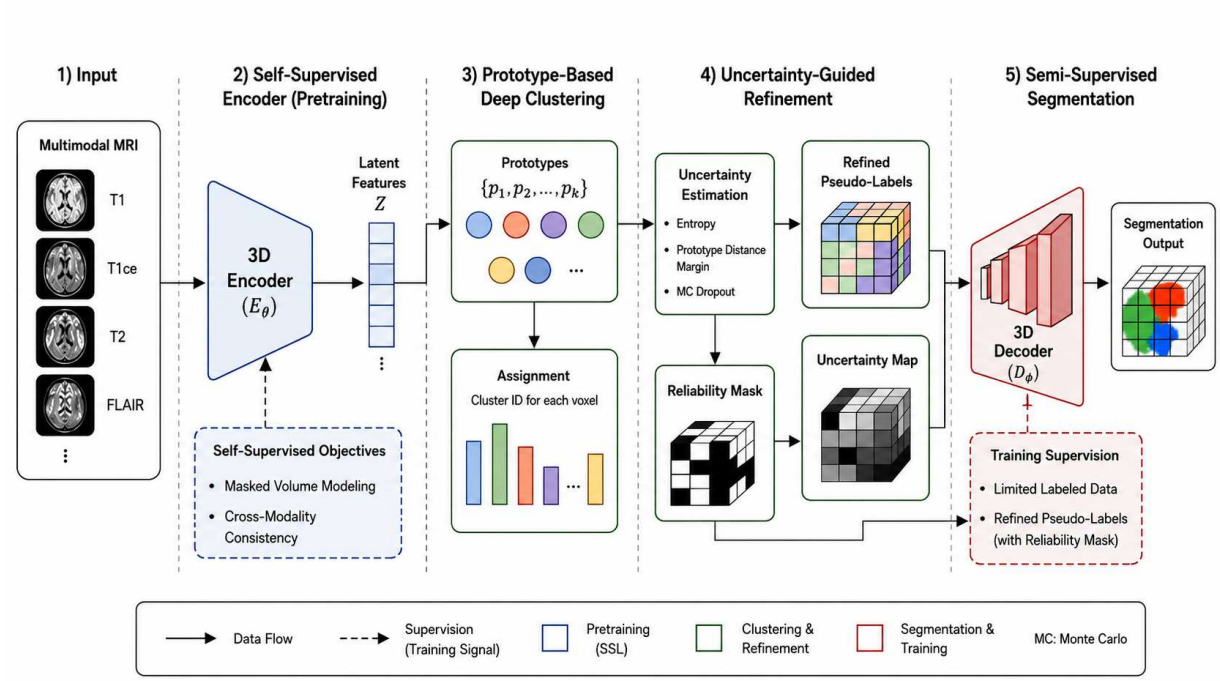


Figure 1: Overall architecture of the proposed UASC-Net framework. Multimodal MRI volumes are first processed by a self-supervised 3D encoder to obtain latent volumetric representations. The learned features are then organized using prototype-based deep clustering to generate initial pseudo-labels. An uncertainty-guided refinement module filters unreliable pseudo-labels through voxel-level reliability estimation. Finally, a semi-supervised decoder is trained using both limited expert annotations and refined pseudo-labels to produce tumor subregion segmentation maps.

Before training, all MRI volumes are processed using a standardized pipeline, consistent with common practice in brain tumor segmentation benchmarks and self-configuring segmentation frameworks (Isensee et al., 2021; Correia de Verdier et al., 2024). First, the modalities of each subject are spatially aligned and resampled to a common resolution when required. Second, intensity normalization is performed independently for each modality to reduce scanner-specific and patient-specific intensity variations. Specifically, z-score normalization is applied to non-background voxels:

$$X_c^{\text{norm}} = \frac{X_c - \mu_c}{\sigma_c + \epsilon}, \tag{3}$$

where X_c is the c-th modality, μ_c and σ_c denote the mean and standard deviation of non-background voxels in that modality, and ϵ is a small constant for numerical stability.

Third, foreground-based cropping is applied to remove irrelevant empty regions and retain the brain region efficiently. During training, 3D patches are extracted from the cropped volumes to reduce memory consumption and increase the number of training samples. The patch extraction strategy is biased toward tumor-containing regions for labeled samples, while random and foreground-guided sampling is used for unlabeled samples. During inference, full-volume segmentation is reconstructed using a sliding-window strategy with overlapping patches. Data augmentation is applied consistently across all modalities and the corresponding masks, including random flipping, random rotation, random cropping, inten-

Table 1: Summary of the main notation used in the proposed method.

Symbol	Description
X	Multimodal MRI input volume
Y	Ground-truth segmentation mask, when available
$\hat{Y}$	Predicted segmentation probability map
$\tilde{Y}$	Refined pseudo-label map
C	Number of input MRI modalities
K	Number of final segmentation classes
D, H, W	Spatial dimensions of the input volume
E_θ	Self-supervised encoder with parameters θ
D_ϕ	Segmentation decoder with parameters ϕ
G_ψ	Projection head for clustering with parameters ψ
Z	Latent feature map generated by the encoder
z_i	Latent feature vector of voxel or patch i
h_i	Projected embedding of z_i in the clustering space
P	Prototype bank
p_q	The q-th learnable prototype
Q	Number of prototypes
a_{iq}	Soft assignment probability of feature i to prototype q
c_i	Hard cluster assignment of feature i
U_i	Final uncertainty score of voxel or patch i
R_i	Binary reliability mask of voxel or patch i
δ	Reliability threshold

sity scaling, Gaussian noise perturbation, and contrast variation. In limited-label experiments, the split between labeled and unlabeled subsets is performed before training to avoid information leakage.

3.3. Self-Supervised Representation Learning

The first learning component of UASC-Net is a self-supervised volumetric encoder E_θ, which learns robust representations from unlabeled multimodal MRI data. Given an input volume X, the encoder produces a latent feature map

$$Z = E_\theta(X), \qquad Z \in \mathbb{R}^{d \times D' \times H' \times W'}, \tag{4}$$

where d denotes the latent feature dimension, and D', H', and W' represent the spatial dimensions of the encoded feature map.

As illustrated in Fig. 3, the encoder follows a hybrid 3D CNN-Transformer design. A 3D convolutional stem first performs patch embedding and low-level feature extraction. The resulting feature maps are then processed through hierarchical encoder stages. Each stage contains convolutional blocks for local spatial modeling and transformer blocks for long-range contextual modeling. This design follows the observation that convolutional operations are effective for local texture extraction, whereas transformer-based modules can model global context in volumetric medical images (Hatamizadeh et al., 2022b,a).

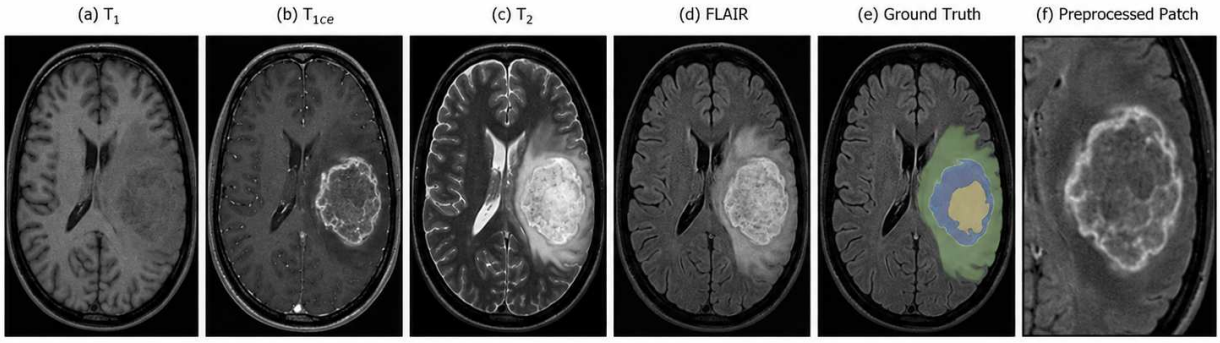


Figure 2: Representative multimodal MRI examples used as input to the proposed framework. The panels show T1, T1ce, T2, and FLAIR images, together with the corresponding ground-truth annotation and a preprocessed tumor-centered patch. The figure highlights the complementary role of different MRI modalities in representing tumor appearance and surrounding abnormal tissue.

Two self-supervised objectives are used to pretrain the encoder. The first objective is masked volume modeling (MVM), inspired by masked image modeling approaches that learn visual representations by reconstructing missing patches (He et al., 2022). Let M denote a binary mask indicating the set of masked patches. The masked input is defined as

$$X^{\text{mask}} = X \odot (1 - M), \tag{5}$$

where $\odot$ denotes element-wise multiplication. The encoder is trained to reconstruct the missing regions:

$$\mathcal{L}_{\text{MVM}} = \frac{1}{|M|} \sum_{i \in M} \left\| \hat{X}_i - X_i \right\|_1, \tag{6}$$

where $\hat{X}_i$ and X_i denote the reconstructed and original voxel or patch values at location i. This objective encourages the encoder to learn anatomical and structural priors from unlabeled MRI volumes.

The second objective is cross-modality consistency (CMC). Since MRI modalities represent complementary tissue properties, representations extracted from spatially corresponding

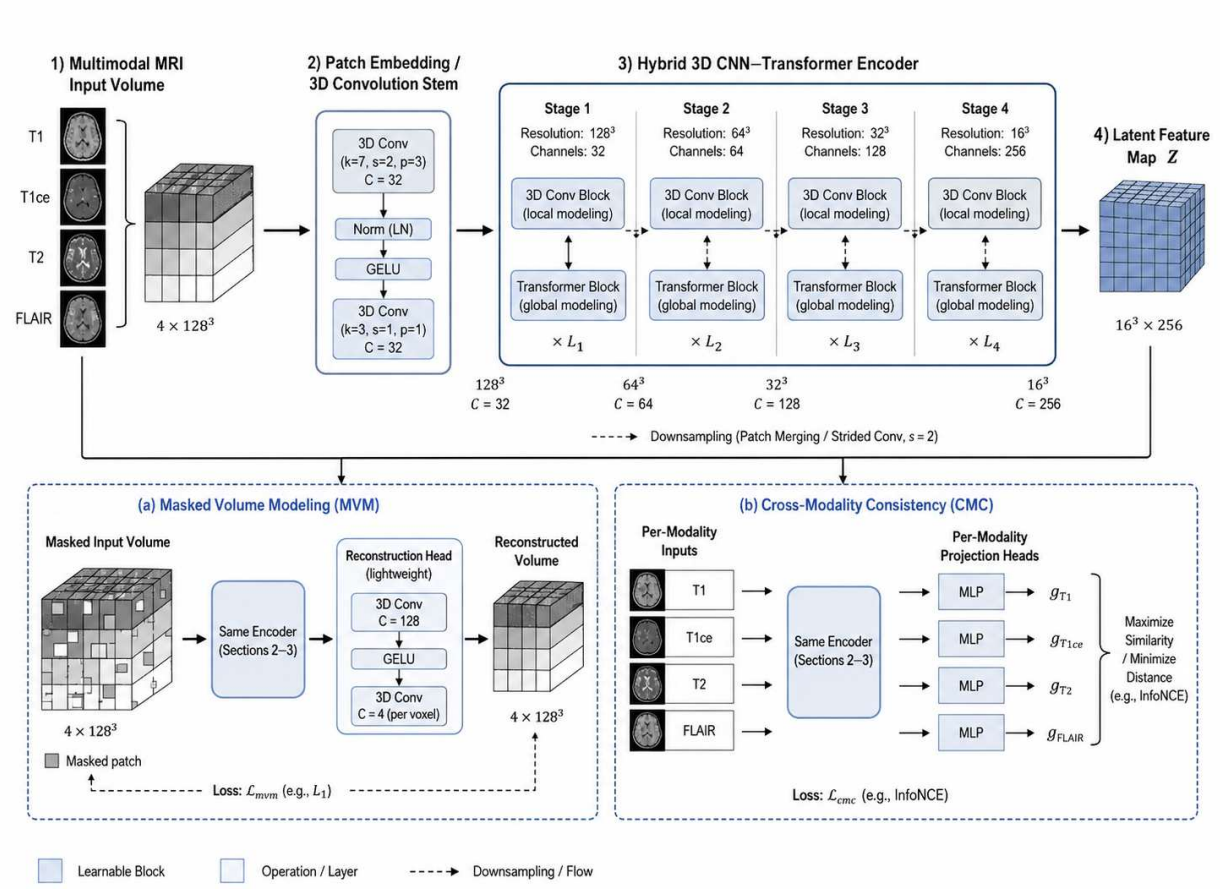


Figure 3: Detailed architecture of the self-supervised representation learning module. The multimodal MRI input is first embedded using a 3D convolutional stem and then processed by a hierarchical hybrid 3D CNN-Transformer encoder. The encoder is pretrained using two complementary objectives: masked volume modeling, which encourages structural representation learning, and cross-modality consistency, which promotes semantically aligned representations across MRI modalities.

regions should remain semantically consistent. Let g_c denote the modality-specific projection of the encoded feature from modality c. For two modalities c and c', the contrastive consistency objective is written as

$$\mathcal{L}_{\mathrm{CMC}} = -\log \frac{\exp(\mathrm{sim}(g_c, g_{c'})/\tau_{\mathrm{cm}})}{\sum_r \exp(\mathrm{sim}(g_c, g_r)/\tau_{\mathrm{cm}})}, \tag{7}$$

where $\mathrm{sim}(\cdot,\cdot)$ denotes cosine similarity and τ_{cm} is a temperature parameter. The overall self-supervised loss is defined as

$$\mathcal{L}_{\mathrm{SSL}} = \mathcal{L}_{\mathrm{MVM}} + \lambda_{\mathrm{cm}}\mathcal{L}_{\mathrm{CMC}}, \tag{8}$$

where λ_{cm} controls the contribution of the cross-modality consistency term.

3.4. Prototype-Based Deep Clustering

After self-supervised pretraining, the latent representation is organized through prototype-based deep clustering. This module is motivated by prior deep clustering and prototype

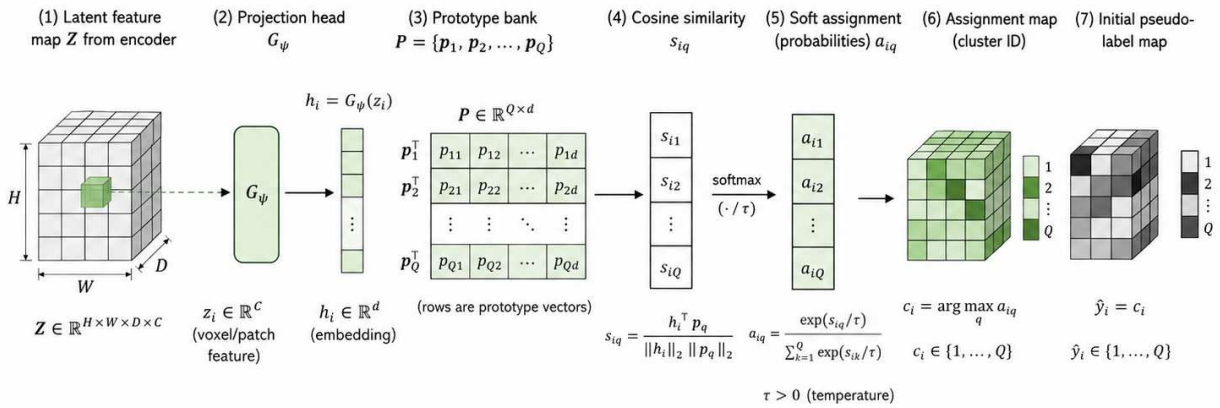


Figure 4: Prototype-based deep clustering module. Latent features extracted by the self-supervised encoder are projected into a clustering embedding space using a projection head. The projected features are compared with a learnable prototype bank using cosine similarity. Temperature-scaled soft assignment probabilities are then computed and converted into hard cluster assignments, which form the initial pseudo-label map for unlabeled MRI volumes.

learning methods, which showed that representation learning can be coupled with cluster assignments or prototypes to discover semantic structure in unlabeled data (Xie et al., 2016; Caron et al., 2018, 2020). The purpose of this stage is to identify recurrent imaging patterns in the latent space and convert them into initial pseudo-labels for unlabeled samples. The complete clustering procedure is shown in Fig. 4.

Let z_i denote the latent feature vector associated with voxel or patch i. A projection head G_ψ maps z_i into a clustering embedding:

$$h_i = G_\psi(z_i), \qquad h_i \in \mathbb{R}^m, \tag{9}$$

where m is the embedding dimension. A learnable prototype bank is defined as

$$P = \{p_1, p_2, \ldots, p_Q\}, \tag{10}$$

where Q denotes the number of prototypes. In practice, Q can be larger than the number of final segmentation classes because a single tumor class may contain multiple visual

subpatterns. For example, edema regions may vary in intensity, extent, and boundary characteristics across patients.

The similarity between the projected feature h_i and the prototype p_q is computed using cosine similarity:

$$s_{iq} = \frac{h_i^\top p_q}{\|h_i\|_2 \|p_q\|_2}. \tag{11}$$

The similarity scores are converted into soft assignment probabilities using a temperature-scaled softmax:

$$a_{iq} = \frac{\exp(s_{iq}/\tau)}{\sum_{r=1}^{Q} \exp(s_{ir}/\tau)}, \tag{12}$$

where τ is the clustering temperature. The hard assignment is then obtained as

$$c_i = \arg\max_q a_{iq}. \tag{13}$$

The initial pseudo-label for voxel or patch i is derived from its prototype assignment:

$$\hat{y}_i^{(0)} = c_i. \tag{14}$$

To encourage compact and discriminative feature-prototype associations, the clustering objective is defined as

$$\mathcal{L}_{\text{cluster}} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{q=1}^{Q} \tilde{a}_{iq} \log(a_{iq}), \tag{15}$$

where $\tilde{a}_{iq}$ is a target assignment distribution, which can be obtained from sharpened assignment probabilities or balanced assignment constraints. The clustering stage transforms the self-supervised latent space into a structured representation suitable for pseudo-label generation.

3.5. Uncertainty-Guided Pseudo-Label Refinement

The pseudo-labels generated by the clustering module are not equally reliable. Ambiguous tumor boundaries, low-contrast regions, heterogeneous lesion appearance, and overlapping clusters may produce unstable assignments. Directly using all pseudo-labels for segmentation training can propagate errors. To address this issue, UASC-Net introduces an uncertainty-guided refinement module that estimates voxel-level reliability before pseudo-labels are used as training supervision. The module is illustrated in Fig. 5. This design is consistent with uncertainty-aware pseudo-labeling strategies that reduce the influence of unreliable pseudo-labels in semi-supervised learning (Gal and Ghahramani, 2016; Rizve et al., 2021; Lu et al., 2023; Chen et al., 2024).

Three complementary uncertainty sources are considered. The first is assignment entropy, which measures the ambiguity of the prototype assignment distribution:

$$U_i^{\text{ent}} = -\sum_{q=1}^{Q} a_{iq} \log(a_{iq}). \tag{16}$$

A high entropy value indicates that voxel or patch i is not confidently assigned to a single prototype.

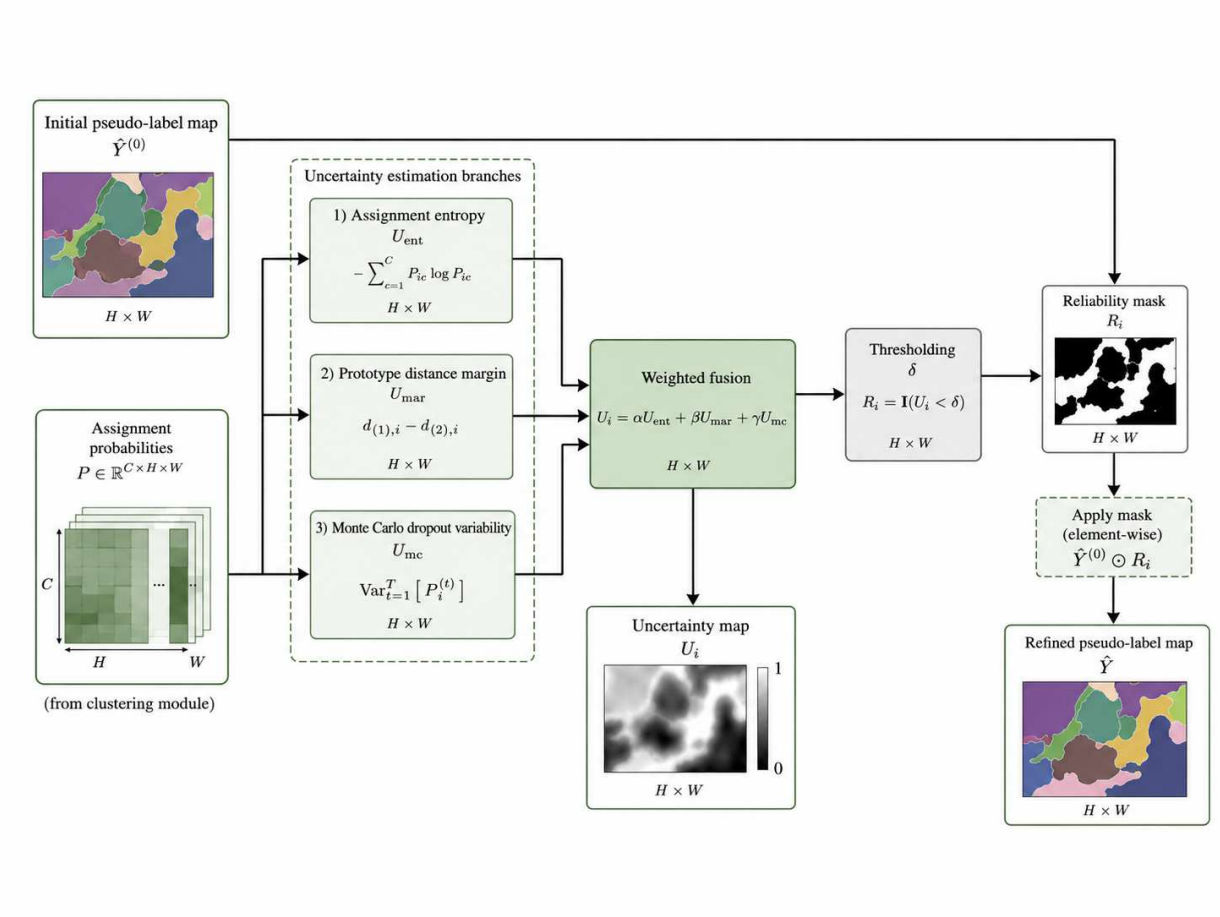


Figure 5: Uncertainty-guided pseudo-label refinement module. Initial pseudo-labels and assignment probabilities from the clustering module are evaluated using three complementary uncertainty estimates: assignment entropy, prototype distance margin, and Monte Carlo dropout variability. These estimates are fused into a voxel-level uncertainty map. A binary reliability mask is then generated by thresholding the uncertainty score and is applied to suppress unreliable pseudo-labels.

The second source is prototype distance margin. Let $d_i^{(1)}$ and $d_i^{(2)}$ denote the distances from feature h_i to its nearest and second-nearest prototypes. The margin-based uncertainty is defined as

$$U_i^{\text{mar}} = \exp\left(-\frac{d_i^{(2)} - d_i^{(1)}}{\tau_m}\right), \tag{17}$$

where τ_m is a scaling temperature. A small margin indicates that the feature is located near a prototype boundary and is therefore less reliable.

The third source is stochastic prediction variability estimated using Monte Carlo dropout (Gal and Ghahramani, 2016). Given T stochastic predictions, the variability term is computed as

$$U_i^{\text{mc}} = \frac{1}{T} \sum_{t=1}^{T} \left\| \hat{y}_i^{(t)} - \bar{y}_i \right\|_2^2, \tag{18}$$

where

$$\bar{y}_i = \frac{1}{T} \sum_{t=1}^{T} \hat{y}_i^{(t)}. \tag{19}$$

The final uncertainty score is defined as a weighted combination of the three uncertainty components:

$$U_i = \alpha U_i^{\text{ent}} + \beta U_i^{\text{mar}} + \gamma U_i^{\text{mc}}, \tag{20}$$

where α, β, and γ are weighting coefficients. Based on this score, a binary reliability mask is generated:

$$R_i = \mathbb{I}(U_i < \delta), \tag{21}$$

where δ is a reliability threshold and $\mathbb{I}(\cdot)$ is the indicator function. The refined pseudo-label is then defined as

$$\tilde{y}_i = R_i \cdot \hat{y}_i^{(0)}. \tag{22}$$

Only reliable pseudo-labels satisfying $R_i = 1$ are used for pseudo-supervised training. This refinement mechanism reduces the influence of noisy cluster assignments and improves the reliability of unlabeled supervision.

3.6. Semi-Supervised Segmentation Decoder and Training Strategy

The final component of UASC-Net is a semi-supervised segmentation decoder. The decoder receives the encoded representation Z and progressively reconstructs a high-resolution voxel-wise segmentation map. As shown in Fig. 6, the decoder follows a U-shaped design with hierarchical upsampling blocks and skip connections from encoder stages, following the localization principle of U-Net-based segmentation models (Ronneberger et al., 2015; Çiçek et al., 2016).

Let the segmentation decoder be denoted as D_ϕ. The predicted segmentation probability map is computed as

$$\hat{Y} = D_\phi(Z), \qquad \hat{Y} \in [0,1]^{K \times D \times H \times W}. \tag{23}$$

The decoder is trained using two complementary supervision sources: limited expert annotations and uncertainty-filtered pseudo-labels.

For labeled samples, the supervised segmentation loss combines Dice loss and cross-entropy loss:

$$\mathcal{L}_{\text{sup}} = \mathcal{L}_{\text{Dice}}(\hat{Y}, Y) + \lambda_{\text{ce}} \mathcal{L}_{\text{CE}}(\hat{Y}, Y), \tag{24}$$

where λ_{ce} controls the contribution of the cross-entropy term. The Dice loss is defined as

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2 \sum_i \hat{y}_i y_i + \epsilon}{\sum_i \hat{y}_i + \sum_i y_i + \epsilon}. \tag{25}$$

For unlabeled samples, the pseudo-supervised loss is computed only over reliable voxels:

$$\mathcal{L}_{\text{pseudo}} = \frac{1}{\sum_i R_i + \epsilon} \sum_i R_i \cdot \mathcal{L}_{\text{CE}}(\hat{y}_i, \tilde{y}_i), \tag{26}$$

where $\tilde{y}_i$ denotes the refined pseudo-label of voxel i. The reliability mask prevents uncertain pseudo-labels from dominating the training process.

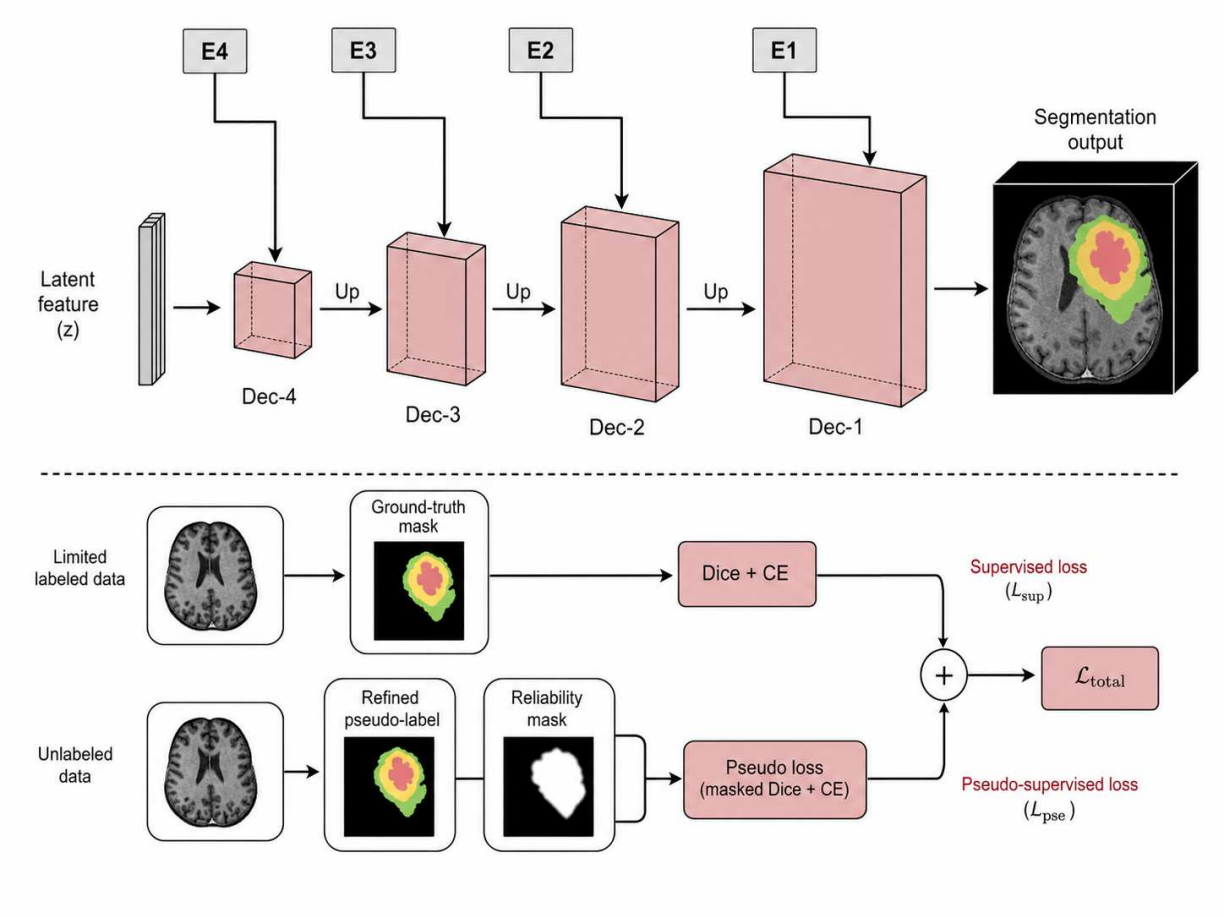


Figure 6: Semi-supervised segmentation decoder and training strategy. The decoder progressively upsamples the latent feature representation and incorporates skip connections from encoder stages to recover spatial details. The training objective combines a supervised loss computed from limited expert annotations and a pseudo-supervised loss computed from uncertainty-filtered pseudo-labels on unlabeled samples.

The total optimization objective is defined as

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{sup}} + \lambda_p \mathcal{L}_{\text{pseudo}} + \lambda_c \mathcal{L}_{\text{cluster}} + \lambda_s \mathcal{L}_{\text{SSL}} + \lambda_u \mathcal{L}_{\text{unc}}, \tag{27}$$

where λ_p, λ_c, λ_s, and λ_u control the contributions of pseudo-supervision, clustering, self-supervised learning, and uncertainty regularization. In practice, $\mathcal{L}_{\text{SSL}}$ is primarily used during encoder pretraining, while $\mathcal{L}_{\text{cluster}}$ and $\mathcal{L}_{\text{pseudo}}$ guide the semi-supervised stage.

3.7. Training Procedure

The complete training procedure of UASC-Net is summarized in Algorithm 1. The training process consists of three major phases. First, the encoder is pretrained using self-supervised objectives on both labeled and unlabeled MRI volumes. Second, prototype-based clustering is performed in the latent space to generate initial pseudo-labels for unlabeled samples. Third, pseudo-labels are refined using uncertainty estimation and then used together with limited expert annotations to train the segmentation decoder.

Algorithm 1 Training procedure of UASC-Net

Require: Labeled dataset $\mathcal{D}_l = \{(X_l, Y_l)\}$, unlabeled dataset $\mathcal{D}_u = \{X_u\}$, prototype bank P, reliability threshold δ

Ensure: Trained encoder E_θ, projection head G_ψ, and segmentation decoder D_ϕ

Preprocess all multimodal MRI volumes and extract 3D patches.
Initialize encoder E_θ, projection head G_ψ, prototype bank P, and decoder D_ϕ.
Pretrain E_θ on $\mathcal{D}_l \cup \mathcal{D}_u$ using $\mathcal{L}_{\text{SSL}}$.
Extract latent features $Z = E_\theta(X)$ for labeled and unlabeled samples.
Project latent features into the clustering space using $h_i = G_\psi(z_i)$.
Compute feature-prototype similarities and soft assignments a_{iq}.
Generate initial pseudo-labels $\hat{y}_i^{(0)}$ using prototype assignments.
Estimate voxel-level uncertainty using entropy, prototype margin, and Monte Carlo dropout variability.
Generate reliability mask $R_i = \mathbb{I}(U_i < \delta)$.
Construct refined pseudo-labels $\tilde{y}_i = R_i \cdot \hat{y}_i^{(0)}$.
Train decoder D_ϕ using expert labels from $\mathcal{D}_l$ and refined pseudo-labels from $\mathcal{D}_u$.
Optimize the overall objective $\mathcal{L}_{\text{total}}$.
Evaluate segmentation performance, clustering quality, pseudo-label reliability, calibration, and ablation variants.

3.8. Implementation-Oriented Architecture Summary

For reproducibility, Table 2 provides a concise implementation-oriented summary of the proposed architecture. The final hyperparameters should be selected using a validation set and reported consistently across all limited-label settings.

4. Experimental Design

4.1. Dataset

The proposed method is intended to be evaluated on the BraTS dataset, which provides multimodal MRI scans and tumor subregion annotations. Each case includes T1, T1ce, T2, and FLAIR modalities. The segmentation targets can be evaluated at both class level and clinically aggregated region level, such as whole tumor, tumor core, and enhancing tumor.

4.2. Limited-Label Evaluation Protocol

To evaluate annotation efficiency, the labeled training set will be subsampled into different label ratios, such as 5%, 10%, 20%, and 100%. The remaining training images will be treated as unlabeled data. This setting allows the proposed method to be compared against supervised and semi-supervised baselines under controlled annotation scarcity.

4.3. Baselines

The following baselines are recommended:

- Fully supervised 3D U-Net.

Table 2: Implementation-oriented configuration of UASC-Net.

Component	Configuration
Input modalities	T1, T1ce, T2, FLAIR
Input representation	3D multimodal volume
Patch size	$128 \times 128 \times 128$ or dataset-dependent
Encoder stem	3D convolutional patch embedding
Encoder backbone	Hybrid 3D CNN-Transformer
Encoder stages	Four hierarchical stages
Feature channels	32, 64, 128, 256
Self-supervised objectives	Masked volume modeling; cross-modality consistency
Projection head	MLP or lightweight convolutional projection head
Number of prototypes	8, 12, or 16
Assignment function	Temperature-scaled softmax over cosine similarities
Uncertainty sources	Entropy, prototype margin, MC dropout variability
Reliability threshold	Validation-based or percentile-based threshold
Decoder	U-shaped 3D decoder with skip connections
Supervised loss	Dice loss + cross-entropy loss
Pseudo-supervised loss	Reliability-masked cross-entropy loss
Training setting	Limited-label semi-supervised learning

- nnU-Net (Isensee et al., 2021).
- Swin UNETR (Hatamizadeh et al., 2022a).
- SSL encoder followed by supervised decoder.
- SSL with pseudo-labeling but without prototype clustering.
- Prototype clustering without uncertainty-guided refinement.
- Full UASC-Net.

4.4. Evaluation Metrics

Segmentation performance will be evaluated using Dice similarity coefficient, Hausdorff distance at the 95th percentile, sensitivity, and specificity. Clustering quality will be assessed using normalized mutual information (NMI), adjusted Rand index (ARI), and cluster purity. Calibration and uncertainty quality will be assessed using expected calibration error (ECE), uncertainty-error correlation, and pseudo-label error reduction before and after refinement.

Table 3: Recommended evaluation metrics for UASC-Net.

Category	Metric	Purpose
Segmentation	Dice score	Region overlap
Segmentation	HD95	Boundary error
Segmentation	Sensitivity / Specificity	Detection reliability
Clustering	NMI	Cluster-label agreement
Clustering	ARI	Pairwise clustering consistency
Pseudo-labeling	Pseudo-label error rate	Reliability of generated labels
Calibration	ECE	Confidence calibration
Uncertainty	Uncertainty-error correlation	Validity of uncertainty maps

4.5. Implementation Details

Table 4: Initial implementation configuration. These values should be finalized after pilot experiments.

Component	Suggested setting
Input modalities	T1, T1ce, T2, FLAIR
Input patch size	$128 \times 128 \times 128$ or dataset-dependent
Encoder	Hybrid 3D CNN–Transformer
Number of prototypes	8, 12, or 16
SSL objectives	Masked volume modeling; cross-modality consistency
Optimizer	AdamW
Learning rate	10^{-4} to 3×10^{-4}
Batch size	GPU-memory dependent
Loss	Dice + CE + pseudo-label loss + clustering loss
Uncertainty threshold	Validation-based or percentile-based

5. Results

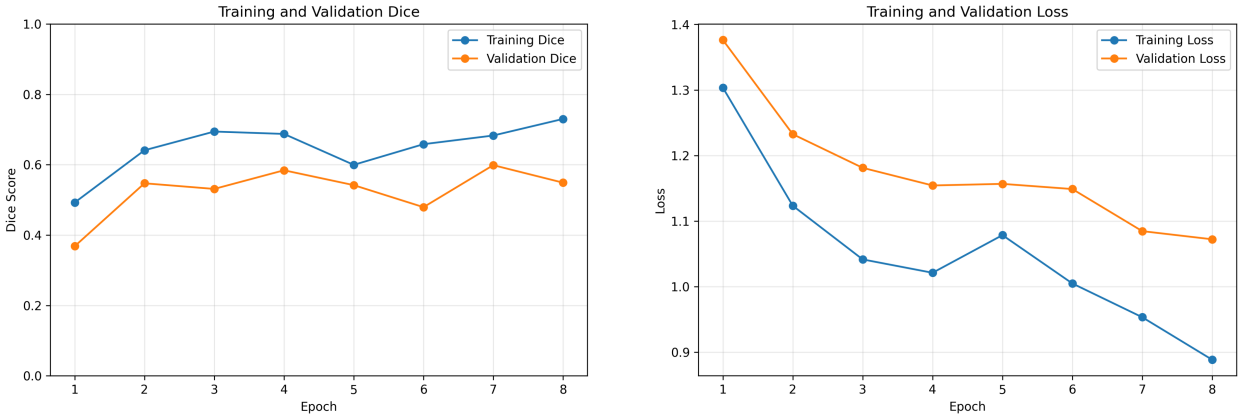
5.1. Evaluation Setting and Reporting Convention

The results reported in this section were obtained on a reproducible pilot subset of the BraTS 2021 Task 1 dataset. Sixteen complete cases were extracted from the official archive; 13 cases were used for training and 3 cases were reserved for validation. All evaluations used $64 \times 64 \times 64$ multimodal 3D patches. The validation split comprised 18 patches, including 12 lesion-containing patches and 6 background-only patches. Both the TinyUNet3D baseline and UASC-Net Lite v2 were trained for eight epochs. The analysis distinguishes between (i) the best validation Dice observed during training using the default threshold of 0.50 and (ii) pooled validation performance after a post-training threshold sweep. Because the threshold was selected on the same validation patches, the calibrated metrics are reported as pilot estimates rather than unbiased test-set estimates.

5.2. Baseline TinyUNet3D Performance

5.2.1. Optimization Dynamics

Figure 7 summarizes the learning behavior of the TinyUNet3D baseline. The training Dice increased from approximately 0.49 to 0.73 over eight epochs, whereas validation Dice rose from 0.37 to a peak of 0.5987 at epoch 7. The corresponding losses decreased over the course of training, indicating stable optimization without numerical divergence. The difference between training and validation trajectories is expected given the small pilot cohort and patch-level sampling variability.



(a) Dice trajectories.

(b) Loss trajectories.

Figure 7: Training and validation behavior of the TinyUNet3D baseline over eight epochs.

5.2.2. Decision-Threshold Calibration

The baseline probability maps were evaluated across thresholds from 0.30 to 0.80. As shown in Figure 8, increasing the threshold reduced sensitivity but substantially improved precision and specificity. The maximum pooled Dice was obtained at $\tau = 0.75$, which was adopted as the baseline operating point. At this threshold, pooled Dice and IoU were 0.7297 and 0.5744, respectively.

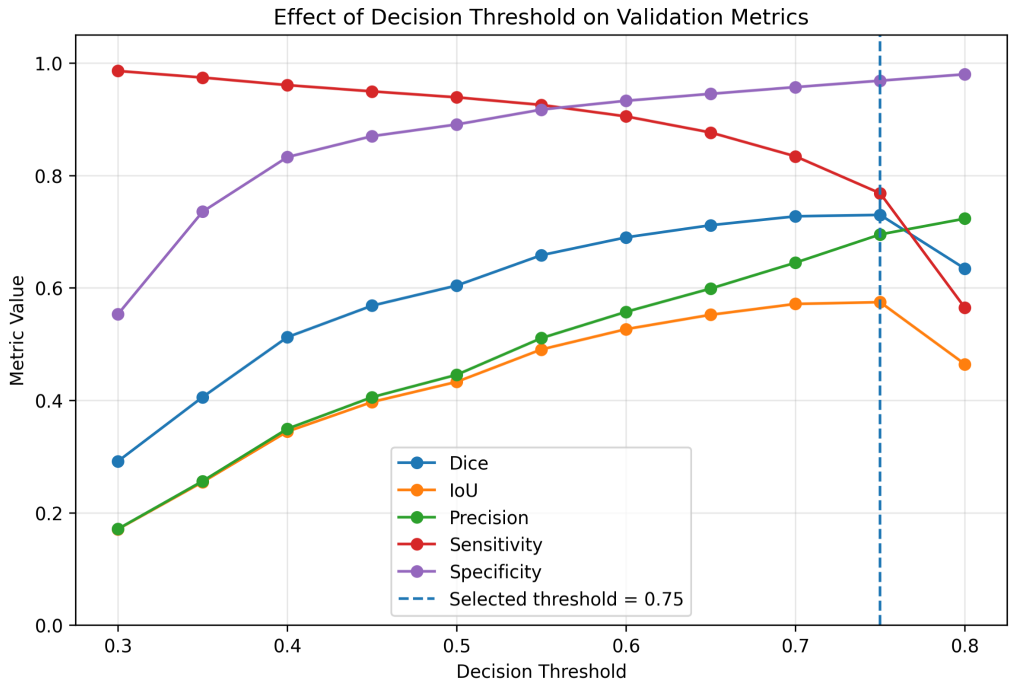


Figure 8: Baseline threshold sweep on validation patches. The selected operating point was $\tau = 0.75$, which maximized pooled Dice.

5.2.3. Baseline Quantitative Results

The calibrated baseline achieved sensitivity of 0.7681, specificity of 0.9685, and precision of 0.6949. The lesion-containing patches had a mean Dice of 0.6831 ± 0.2595. In contrast, the six background-only patches contained a mean of 27,579.7 predicted tumor voxels, indicating a non-negligible false-positive burden. This distinction is important: the high pooled specificity does not eliminate false positives at the patch level, particularly when a model is applied to a small and heterogeneous validation cohort.

5.2.4. Representative Baseline Prediction

Figure 9 shows a representative baseline prediction after calibration. The five panels contain the FLAIR input, reference mask, thresholded output, tumor-probability map, and an overlay. The baseline identifies the dominant lesion component and yields a coherent probability field, while the overlay reveals localized boundary disagreement. Only one baseline visualization is included here to avoid duplicating the same prediction under two different filenames.

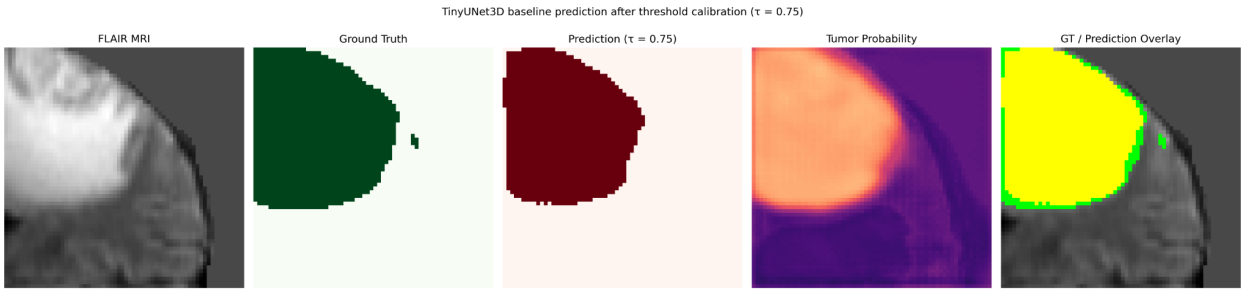
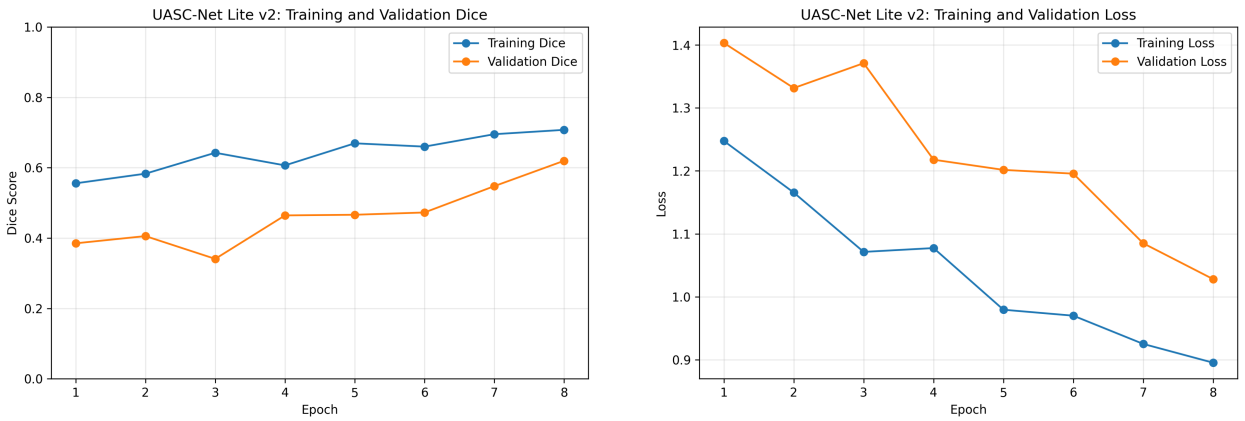


Figure 9: Representative TinyUNet3D prediction after threshold calibration ($\tau = 0.75$). Yellow indicates overlap between ground truth and prediction; red and green denote prediction-only and ground-truth-only regions, respectively.

5.3. UASC-Net Lite v2 Performance

5.3.1. Optimization and Validation Behavior

UASC-Net Lite v2 showed stable learning over eight epochs (Figure 10). Training Dice increased from 0.5556 to 0.7076, while validation Dice increased from 0.3850 to 0.6196. The loss curves decreased throughout training, with the final validation loss reaching 1.0281. The best checkpoint was therefore selected at epoch 8 under the default threshold of 0.50.



(a) Dice trajectories.

(b) Loss trajectories.

Figure 10: Training and validation behavior of UASC-Net Lite v2.

5.3.2. Uncertainty Learning

The uncertainty branch also converged over the training period (Figure 11). Validation uncertainty loss decreased from 0.1043 at the first epoch to 0.0487 at epoch 8. This result indicates that the auxiliary branch was optimized jointly with segmentation, although it should not be interpreted as a formal calibration guarantee. In the representative qualitative example described below, the mean uncertainty value was 0.2615.

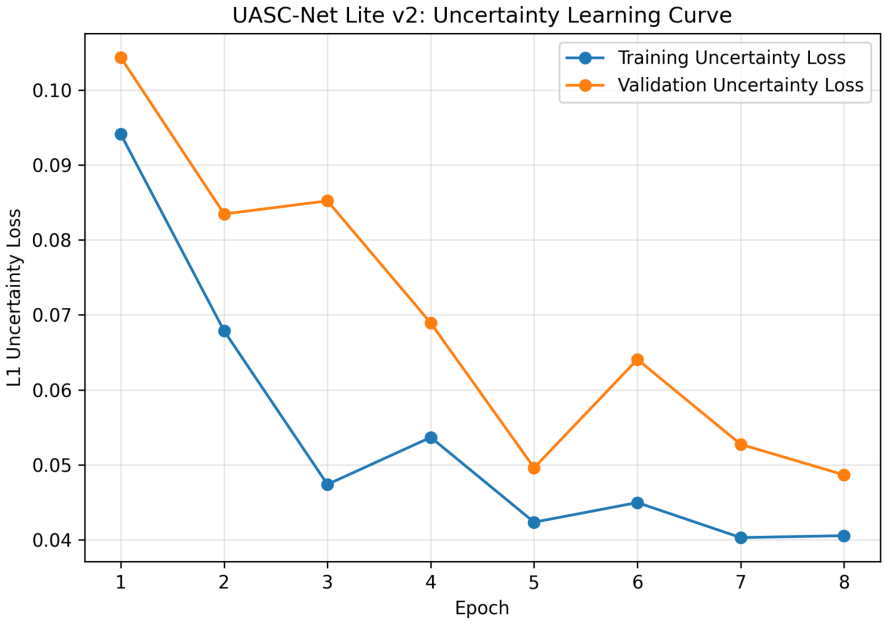


Figure 11: Training and validation uncertainty losses for UASC-Net Lite v2.

5.3.3. Threshold-Calibrated Performance

The UASC-Net Lite v2 probability maps were evaluated at thresholds from 0.30 to 0.80. The optimum pooled Dice was obtained at $\tau = 0.80$. At this operating point, the model achieved Dice of 0.7381, IoU of 0.5849, sensitivity of 0.8609, specificity of 0.9532, and precision of 0.6460. The increase in the selected threshold relative to the baseline (0.80 versus 0.75) is consistent with a more conservative conversion of the probability field into a binary mask.

5.4. Comparison with the Baseline

Table 5 reports the main pilot results. UASC-Net Lite v2 improved the calibrated pooled Dice from 0.7297 to 0.7381, an absolute gain of 0.0084, and improved IoU from 0.5744 to 0.5849. The largest change was observed in sensitivity, which increased from 0.7681 to 0.8609. This improved lesion recall was accompanied by lower precision (0.6460 versus 0.6949) and slightly lower specificity (0.9532 versus 0.9685), reflecting a clinically relevant recall–false-positive trade-off rather than an across-the-board improvement in every metric.

Table 5: Pilot comparison of TinyUNet3D and UASC-Net Lite v2 after validation-set threshold calibration.

Metric	TinyUNet3D	UASC-Net Lite v2
Training cases	13	13
Validation cases	3	3
Validation patches	18	18
Patch size	$64 \times 64 \times 64$	$64 \times 64 \times 64$
Epochs	8	8
Best validation Dice at $\tau = 0.50$	0.5987 (epoch 7)	**0.6196 (epoch 8)**
Selected threshold	0.75	0.80
Pooled Dice	0.7297	**0.7381**
Pooled IoU	0.5744	**0.5849**
Sensitivity	0.7681	**0.8609**
Specificity	**0.9685**	0.9532
Precision	**0.6949**	0.6460

Figure 12 visualizes the same comparison. The proposed model improved overlap and sensitivity, whereas the baseline retained an advantage in precision and specificity. Thus, the results support the interpretation that uncertainty-aware modeling improved lesion coverage, but still requires additional false-positive control before any clinical use.

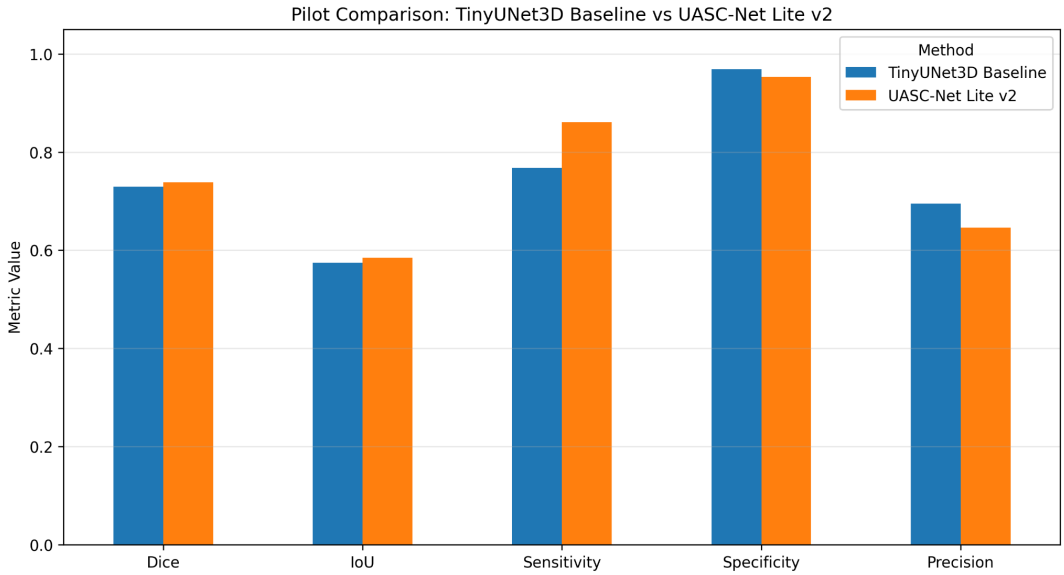


Figure 12: Metric-level comparison between the TinyUNet3D baseline and UASC-Net Lite v2 on the pilot validation subset.

5.5. Qualitative Comparison, Uncertainty, and Attention

Figure 13 compares both models on one representative validation slice. The figure contains the FLAIR image, ground-truth mask, baseline prediction and overlay, UASC-Net prediction and overlay, as well as the UASC uncertainty and attention maps. On this slice, the baseline Dice was 0.7867 and the UASC-Net Dice was 0.7943. The example is consistent

with the pooled results, but it is not a substitute for them because a single slice cannot characterize overall model behavior.

The UASC uncertainty map highlights regions in which the confidence estimate varies spatially, while the attention map identifies regions emphasized by the auxiliary attention mechanism. These maps are intended as interpretability outputs rather than independent clinical evidence. Their apparent alignment with lesion-related structures should be examined in larger studies with calibration metrics and expert review.

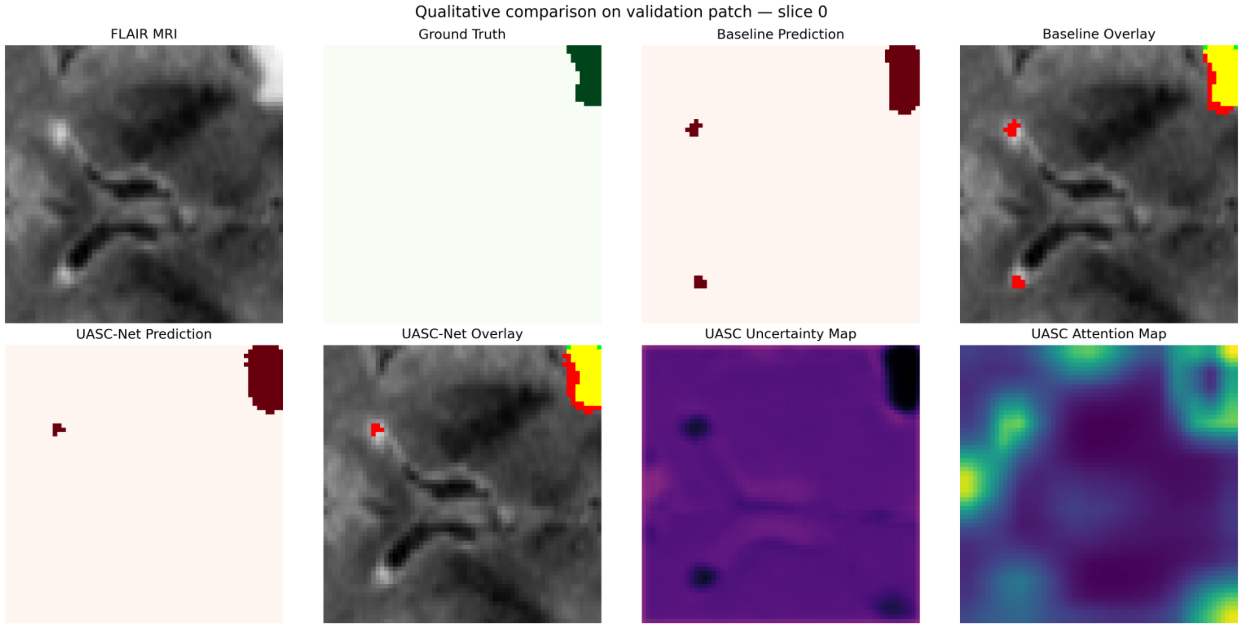


Figure 13: Qualitative comparison on a representative validation slice. Top row: FLAIR MRI, ground truth, TinyUNet3D prediction, and baseline overlay. Bottom row: UASC-Net prediction, UASC-Net overlay, uncertainty map, and attention map.

5.6. Summary of Findings

The pilot experiments yield three main findings. First, UASC-Net Lite v2 produced modest but consistent improvements in overlap-based metrics compared with the compact 3D U-Net baseline. Second, the principal numerical advantage was higher sensitivity, indicating improved recovery of tumor voxels, while precision and specificity decreased modestly. Third, the uncertainty and attention branches added model-inspection outputs that are absent from the baseline. These observations motivate broader evaluation, but the small cohort, patch-level validation, and validation-set threshold selection prevent claims of clinical superiority or generalization beyond the present pilot setting.

6. Discussion

The pilot comparison suggests that incorporating uncertainty-aware and auxiliary representation components into a compact 3D segmentation framework can shift the operating behavior toward higher lesion recall. The proposed UASC-Net Lite v2 improved pooled Dice by 0.0084 and increased sensitivity by 0.0928 relative to the TinyUNet3D baseline after

threshold calibration. In a brain tumor segmentation setting, this change can be valuable because missed tumor voxels may be clinically consequential. However, the gain must be interpreted jointly with the reduction in precision and specificity. The proposed model produced more positive predictions, which recovered additional tumor voxels but also introduced more false-positive voxels.

The uncertainty branch is useful primarily as an interpretability mechanism in the present experiment. Its training and validation losses decreased over time, and the qualitative map showed spatially varying confidence. Nonetheless, uncertainty quality was not established by a formal calibration analysis such as expected calibration error, risk–coverage curves, or correlation with voxel-level error. Similarly, the prototype and attention outputs should be regarded as intermediate explanatory signals rather than validated biological markers.

Several limitations require emphasis. First, only 16 cases were used, with three validation cases and 18 extracted validation patches. Second, the segmentation target was binary tumor versus background; the experiment did not report the conventional BraTS WT, TC, and ET subregion metrics. Third, the decision threshold was chosen using the validation patches, which can introduce optimistic bias. Fourth, patch-level observations are not equivalent to full-volume patient-level evaluation. Finally, the architecture presented in this paper is a Lite pilot instantiation of the broader UASC-Net concept; comprehensive self-supervised pretraining, prototype ablation, calibration analysis, and limited-label benchmarking remain future work.

Future evaluation should use patient-level cross-validation or an independent test set, full-volume inference with overlap-aware patch aggregation, multiclass BraTS labels, and direct comparison with stronger baselines such as nnU-Net and transformer-based 3D segmentation models. Ablation experiments should quantify the independent contribution of the uncertainty branch, attention mechanism, and prototype representation. These steps are required before drawing conclusions about clinical utility or state-of-the-art performance.

7. Conclusion

This work presented a reproducible pilot evaluation of UASC-Net Lite v2 for multimodal MRI brain tumor segmentation. On a 16-case BraTS 2021 subset, the proposed model achieved a threshold-calibrated pooled Dice of 0.7381 and IoU of 0.5849, exceeding the corresponding TinyUNet3D baseline values of 0.7297 and 0.5744. The principal improvement was increased sensitivity (0.8609 versus 0.7681), accompanied by lower precision and specificity. UASC-Net Lite v2 additionally provided uncertainty and attention maps that facilitate inspection of model behavior. These results support further development of uncertainty-aware segmentation under limited-data conditions, but remain pilot findings due to the small sample size, patch-level evaluation, and validation-set threshold selection. Larger patient-level studies with multiclass targets, external validation, and dedicated calibration analyses are needed to establish robustness and clinical relevance.

Ethics Statement

This study involved a secondary analysis of publicly available, de-identified MRI data from the BraTS 2021 Task 1 dataset. The authors did not recruit human participants, collect new biological samples, access identifiable personal information, or intervene in patient care.

All procedures related to the original data collection, de-identification, and data sharing were governed by the BraTS dataset providers and the applicable dataset-access terms. No new institutional ethics approval was required for this secondary analysis of publicly available, de-identified data. Informed consent from individual participants was not applicable to the present study because the authors had no direct interaction with participants and no access to identifiable personal information.

All methods were performed in accordance with relevant guidelines and regulations, as well as the applicable BraTS and Kaggle data-use conditions.

Data Availability

The data used in this study were obtained from the publicly available BraTS 2021 Task 1 dataset: https://www.kaggle.com/datasets/dschettler8845/brats-2021-task1

The dataset contains de-identified multimodal brain MRI data and is accessible subject to the applicable Kaggle and BraTS dataset-access terms. The original MRI data are not redistributed by the authors. Source code, experimental configuration files, and reproducibility materials for the UASC-Net Lite v2 pilot implementation are publicly available at: https://github.com/Mojtaba-jahanian/UASC-Net

Code Availability

The source code for the reproducible UASC-Net Lite v2 pilot implementation is publicly available at:

https://github.com/Mojtaba-jahanian/UASC-Net

The repository includes the TinyUNet3D baseline, UASC-Net Lite v2 implementation, BraTS-style multimodal MRI data loading, tumor-aware 3D patch sampling, training and evaluation scripts, threshold-sweep utilities, configuration files, and reproducibility documentation. The repository reproduces the pilot implementation reported in this study and does not claim to implement the complete conceptual self-supervised, deep-clustering, pseudo-label refinement, or semi-supervised framework. MRI data are not redistributed and must be obtained separately under the applicable BraTS and Kaggle terms.

Funding

This research received no specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors acknowledge the organizers and contributors of the BraTS benchmark for making the dataset available for research use.

References

Caron, M., Bojanowski, P., Joulin, A., Douze, M., 2018. Deep clustering for unsupervised learning of visual features, in: Proceedings of the European Conference on Computer Vision, pp. 132–149.

Caron, M., Misra, I., Mairal, J., Goyal, P., Bojanowski, P., Joulin, A., 2020. Unsupervised learning of visual features by contrasting cluster assignments, in: Advances in Neural Information Processing Systems, pp. 9912–9924.

Chen, T., Kornblith, S., Norouzi, M., Hinton, G., 2020. A simple framework for contrastive learning of visual representations, in: Proceedings of the 37th International Conference on Machine Learning, PMLR. pp. 1597–1607.

Chen, Y., et al., 2024. Evidence-based uncertainty-aware semi-supervised medical image segmentation. Computers in Biology and Medicine 171, 108079. doi:10.1016/j.compbiomed.2024.108079.

Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O., 2016. 3d u-net: Learning dense volumetric segmentation from sparse annotation, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016, Springer. pp. 424–432. doi:10.1007/978-3-319-46723-8_49.

Gal, Y., Ghahramani, Z., 2016. Dropout as a bayesian approximation: Representing model uncertainty in deep learning, in: Proceedings of the 33rd International Conference on Machine Learning, PMLR. pp. 1050–1059.

Haghighi, F., Hosseinzadeh Taher, M.R., Gotway, M.B., Liang, J., 2024. Self-supervised learning for medical image analysis: Discriminative, restorative, or adversarial? Medical Image Analysis 94, 103086. doi:10.1016/j.media.2024.103086.

Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D., 2022a. Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images, in: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, Springer. pp. 272–284. doi:10.1007/978-3-031-08999-2_22.

Hatamizadeh, A., Yang, D., Roth, H., Xu, D., 2022b. Unetr: Transformers for 3d medical image segmentation, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 574–584.

He, K., Chen, X., Xie, S., Li, Y., Dollár, P., Girshick, R., 2022. Masked autoencoders are scalable vision learners, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 16000–16009.

Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H., 2021. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods 18, 203–211. doi:10.1038/s41592-020-01008-z.

Liu, Z., et al., 2024. A review of self-supervised, generative, and few-shot deep learning methods for medical image segmentation. Medical Image Analysis 95, 103187. doi:10.1016/j.media.2024.103187.

Lu, L., Yin, M., Fu, L., Yang, F., 2023. Uncertainty-aware pseudo-label and consistency for semi-supervised medical image segmentation. Biomedical Signal Processing and Control 79, 104203. doi:10.1016/j.bspc.2022.104203.

Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al., 2015. The multimodal brain tumor image segmentation benchmark (brats). IEEE Transactions on Medical Imaging 34, 1993–2024. doi:10.1109/TMI.2014.2377694.

Rizve, M.N., Duarte, K., Rawat, Y.S., Shah, M., 2021. In defense of pseudo-labeling: An uncertainty-aware pseudo-label selection framework for semi-supervised learning, in: International Conference on Learning Representations, pp. 1–20.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Springer. pp. 234–241. doi:10.1007/978-3-319-24574-4_28.

Shurrab, S., Duwairi, R., 2022. Self-supervised learning methods and applications in medical imaging analysis: a survey. PeerJ Computer Science 8, e1045. doi:10.7717/peerj-cs.1045.

Tarvainen, A., Valpola, H., 2017. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results, in: Advances in Neural Information Processing Systems, pp. 1195–1204.

Correia de Verdier, M., Saluja, R., Gagnon, L., LaBella, D., Baid, U., et al., 2024. The 2024 brain tumor segmentation (brats) challenge: Glioma segmentation on post-treatment mri. arXiv preprint arXiv:2405.18368 .

Xie, J., Girshick, R., Farhadi, A., 2016. Unsupervised deep embedding for clustering analysis, in: Proceedings of the 33rd International Conference on Machine Learning, PMLR. pp. 478–487.

Zeng, X., et al., 2024. Self-supervised learning framework application for medical image analysis: a review and summary. BioMedical Engineering OnLine 23, 107. doi:10.1186/s12938-024-01299-9.