IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p-ISSN: 2278-8727, Volume 28, Issue 3, Ser. II (May. – June. 2026), PP 55-64
www.iosrjournals.org

Clinical Missingness-Aware mmFormer for Incomplete Multimodal Brain Tumor Segmentation

Thanh Trung Giang[1]
[1](School Of Information and Communications Technology, Hanoi University of Industry, Hanoi, Vietnam.)

Abstract: *Multimodal magnetic resonance imaging plays an important role in brain tumor segmentation because different MRI sequences provide complementary information about tumor subregions; however, most existing multimodal segmentation methods assume that all modalities are available and reliable during inference, an assumption that is often unrealistic in clinical practice where MRI modalities may be missing or degraded by noise, motion artifacts, blur, or low signal quality. To address this limitation, this study proposes Clinical Missingness-Aware mmFormer (CMA-mmFormer) for incomplete multimodal brain tumor segmentation. The proposed framework extends mmFormer by introducing a clinical missingness simulation protocol, a Missingness Encoder, and a Reliability Gating mechanism. The model was evaluated on the BraTS 2018 dataset using four MRI modalities (FLAIR, T1ce, T1, and T2) under multiple missingness scenarios, including complete modalities, missing T1ce, FLAIR + T2 limited protocol, single-modality input, and quality-degraded modalities. Segmentation performance was assessed using Dice score and HD95, while prediction reliability was evaluated using expected calibration error, Brier score, negative log-likelihood, and risk-coverage analysis. Experimental results showed that CMA-mmFormer outperformed the original mmFormer baseline, improving WT Dice from 84.20% to 87.10%, TC Dice from 76.80% to 81.50%, and ET Dice from 72.40% to 78.30%, while also reducing HD95 values across all tumor subregions. Under clinical missingness scenarios, CMA-mmFormer demonstrated consistent improvements, particularly in missing T1ce and quality-degraded settings. In addition, the model achieved better calibration, reducing ECE from 0.082 to 0.049 and AURC from 0.214 to 0.151. These findings indicate that CMA-mmFormer improves incomplete multimodal brain tumor segmentation by explicitly modeling both modality availability and modality reliability, and that the proposed Missingness Encoder and Reliability Gating mechanism enhance segmentation robustness and prediction reliability under clinically realistic missing and degraded MRI conditions, highlighting the potential of clinical missingness-aware modeling for trustworthy multimodal medical image analysis.*
Key Word*: Brain tumor segmentation; multimodal MRI; missing modality; mmFormer; clinical missingness; Reliability Gating; BraTS 2018.*

Date of Submission: 02-06-2026 Date of Acceptance: 13-06-2026

I. Introduction

Brain tumor segmentation from magnetic resonance imaging (MRI) plays a crucial role in modern neuro-oncology because accurate delineation of tumor regions supports diagnosis, surgical planning, radiotherapy guidance, treatment monitoring, and longitudinal assessment of disease progression. Gliomas, one of the most common and aggressive primary brain tumors, exhibit substantial heterogeneity in shape, size, and appearance, making automated segmentation a challenging task. Multimodal MRI provides complementary information through fluid-attenuated inversion recovery (FLAIR), T1-weighted (T1), contrast-enhanced T1-weighted (T1ce), and T2-weighted (T2) sequences, enabling the identification of edema, necrotic tissue, non-enhancing tumor, and enhancing tumor regions. To accelerate research in this area, the Brain Tumor Segmentation (BraTS) challenges have established standardized datasets, expert annotations, and evaluation protocols that serve as the most widely used benchmark for brain tumor segmentation studies [1], [2].

Recent advances in deep learning have significantly improved the performance of automated brain tumor segmentation systems. Early convolutional neural network (CNN)-based methods demonstrated superior feature learning capabilities compared with traditional handcrafted approaches [3], [11], [12]. Encoder–decoder architectures such as U-Net and its variants became the dominant framework for biomedical image segmentation because of their ability to integrate local and contextual information effectively [13], [14]. Subsequent developments incorporated three-dimensional convolutions, residual connections, dense feature propagation, and attention mechanisms to better capture volumetric structures and long-range dependencies within MRI data [15]–[18]. More recently, transformer-based architectures have gained increasing attention due to their ability to model global contextual relationships through self-attention mechanisms, leading to improved segmentation performance in complex medical imaging tasks [4], [19]–[21].

Despite these advances, most multimodal segmentation models assume that all MRI modalities are available during both training and inference. In clinical practice, however, this assumption is frequently violated. Differences in acquisition protocols, scanner availability, examination duration, patient tolerance, institutional policies regarding contrast-agent administration, and economic constraints can result in incomplete MRI examinations. Missing modalities may also arise from corrupted acquisitions, registration failures, incomplete archives, or quality-control exclusions. As a result, models trained on complete multimodal datasets often experience substantial performance degradation when deployed in real-world settings where one or more modalities are unavailable [22]–[24].

To improve robustness under incomplete multimodal conditions, several strategies have been proposed. Sequence dropout and hetero-modal learning approaches aim to reduce dependence on specific MRI sequences by exposing models to missing modalities during training [5]. Other studies have explored modality synthesis and image translation techniques based on generative adversarial networks to reconstruct unavailable modalities before segmentation [25]–[27]. Additional methods have investigated modality-invariant feature learning, latent-space completion, knowledge distillation, and modality-aware fusion mechanisms to preserve segmentation performance despite incomplete inputs [6], [7], [28]–[30]. Among transformer-based approaches, mmFormer has emerged as a representative framework by combining modality-specific encoders with intra-modal and inter-modal transformers to learn robust multimodal representations under missing-modality conditions [8].

Although existing methods have improved robustness to missing modalities, most approaches treat missingness as a binary condition in which a modality is either present or absent. This assumption does not fully reflect clinical reality. In practice, missingness patterns are often influenced by acquisition protocols, institutional workflows, patient-specific factors, and diagnostic requirements. Furthermore, a modality may be available but exhibit reduced diagnostic quality due to motion artifacts, noise, low signal-to-noise ratio, susceptibility distortions, intensity inhomogeneity, or inadequate contrast enhancement. Such variations can introduce uncertainty into multimodal fusion and negatively affect segmentation performance when all available modalities are treated equally. Therefore, there is a need for segmentation frameworks that consider both modality availability and modality reliability during feature integration.

Another important challenge concerns the reliability of model predictions. Most brain tumor segmentation studies focus primarily on overlap-based and boundary-based metrics such as Dice score and Hausdorff distance, which measure segmentation accuracy but provide limited information about prediction confidence. In safety-critical medical applications, reliable uncertainty estimation and confidence calibration are essential because overconfident incorrect predictions may adversely influence clinical decision-making. Recent studies have highlighted the importance of calibration metrics such as Expected Calibration Error (ECE), Brier Score, and Negative Log-Likelihood (NLL) for evaluating the trustworthiness of deep learning systems in healthcare [9], [10], [31], [32].

Motivated by these challenges, this study proposes a Clinical Missingness-Aware mmFormer (CMA-mmFormer) for incomplete multimodal brain tumor segmentation. The proposed framework extends mmFormer through a clinical missingness simulation protocol, a Missingness Encoder, and a Reliability Gating mechanism that jointly model modality availability and reliability. The method is evaluated on the BraTS 2018 dataset under multiple clinically motivated missingness scenarios using both segmentation accuracy metrics and reliability-oriented evaluation measures. The main contributions of this work are: (1) a clinical missingness simulation protocol that reflects realistic MRI acquisition conditions; (2) a missingness-aware representation module that jointly encodes modality availability and reliability; (3) a reliability gating mechanism for adaptive multimodal feature fusion; and (4) a comprehensive evaluation framework combining segmentation, calibration, and uncertainty metrics.

II. Material And Methods

Dataset and preprocessing

This study used the BraTS 2018 dataset for multimodal brain tumor segmentation. BraTS 2018 provides pre-operative multimodal MRI scans of glioma patients acquired from multiple institutions using different clinical protocols and scanners. Each case includes four structural MRI modalities: native T1-weighted image, contrast-enhanced T1-weighted image (T1ce), T2-weighted image, and FLAIR image [1], [2]. The ground-truth annotations include the necrotic and non-enhancing tumor core (NCR/NET), peritumoral edema, and enhancing tumor regions. In the original BraTS annotation protocol, the labels are defined as background, NCR/NET, edema, and enhancing tumor, corresponding to labels 0, 1, 2, and 4, respectively [1], [2].

In this study, the MRI modalities were arranged in the following order throughout preprocessing, training, and evaluation: FLAIR, T1ce, T1, and T2. To make the segmentation labels compatible with the network output channels, the original enhancing tumor label 4 was remapped to label 3. Therefore, the final label set used in the experiments was defined as follows: 0 for background, 1 for NCR/NET, 2 for edema, and 3

for enhancing tumor. Based on these labels, the commonly used tumor sub-regions were defined as whole tumor (WT: labels 1, 2, and 3), tumor core (TC: labels 1 and 3), and enhancing tumor (ET: label 3).

The BraTS 2018 images were already provided after standard preprocessing, including co-registration to the same anatomical template, interpolation to isotropic 1mm³ resolution, and skull stripping [1], [2]. In the present implementation, all NIfTI volumes were converted into NumPy array format for efficient loading during training. Each case was stored as a multi-channel volume corresponding to the four MRI modalities. The converted data were cached to avoid repeated preprocessing in later training sessions.

To ensure reproducibility, the data split followed the same logic as the mmFormer baseline implementation. All case files were first sorted by case name, and then shuffled using a fixed random seed. The dataset was divided into three folds, where fold 0 was used for testing and the remaining folds were used for training. This split strategy was adopted to maintain consistency with the baseline setting and to allow fair comparison between mmFormer and the proposed CMA-mmFormer model.

For missing-modality experiments, modality availability was represented by a binary mask following the same modality order: FLAIR, T1ce, T1, and T2. A value of 1 indicated that the corresponding modality was available, whereas a value of 0 indicated that the modality was missing. For example, the mask [1, 1, 1, 1] represented the complete four-modality setting, while [1, 0, 0, 1] represented a limited acquisition protocol using only FLAIR and T2. These masks were later used in the clinical missingness simulation protocol and in the proposed missingness-aware fusion mechanism.

Baseline model

The baseline model used in this study was mmFormer, a multimodal medical transformer proposed for incomplete multimodal brain tumor segmentation [8]. mmFormer was selected as the baseline because it was specifically designed for brain tumor segmentation from incomplete multimodal MRI data and was evaluated on the BraTS 2018 dataset. Unlike conventional multimodal segmentation models that assume all modalities are available, mmFormer aims to segment tumor regions when any non-empty subset of MRI modalities is available.

The mmFormer architecture consists of three main components: modality-specific encoders, transformer-based multimodal fusion, and a segmentation decoder. The modality-specific encoders are used to extract features from individual MRI modalities. These encoders combine convolutional feature extraction with transformer-based modeling so that both local image patterns and global contextual information can be represented. This design is important for brain tumor segmentation because tumor regions may vary substantially in size, shape, intensity, and spatial distribution.

After modality-specific feature extraction, mmFormer uses an inter-modal transformer to model relationships across MRI modalities. This component is designed to learn modality-invariant representations by capturing long-range correlations among features extracted from different modalities. When one or more modalities are missing, the model uses modality availability information to guide feature fusion and reduce the negative effect of incomplete inputs. The fused multimodal representation is then passed to the decoder, which progressively upsamples the features and produces the final segmentation map.

In this study, mmFormer was used as the primary baseline for two reasons. First, it is directly related to the research problem of incomplete multimodal brain tumor segmentation. Second, its architecture provides a suitable foundation for the proposed CMA-mmFormer, which extends the original framework by introducing clinical missingness simulation, a Missingness Encoder, and Reliability Gating. Therefore, the comparison between mmFormer and CMA-mmFormer allows the effect of clinical missingness-aware modeling and reliability-aware fusion to be evaluated under the same dataset and experimental setting.

Proposed method

This study proposes Clinical Missingness-Aware mmFormer (CMA-mmFormer) for incomplete multimodal brain tumor segmentation. The proposed model is developed by extending the original mmFormer framework [8] to better handle clinically realistic missing-modality conditions. While mmFormer mainly focuses on modality availability, CMA-mmFormer further considers modality reliability. This distinction is important because, in clinical MRI, a modality may not only be completely missing but may also be available with degraded quality due to motion artifacts, noise, blur, or low signal quality.

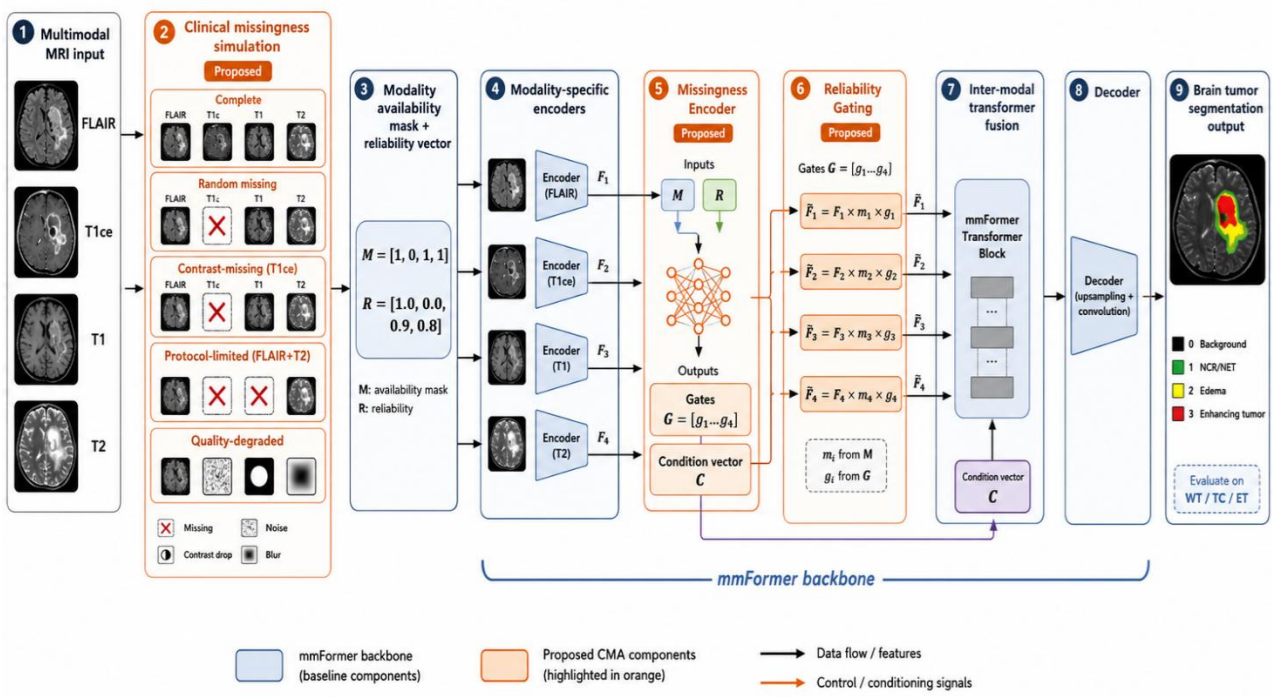

Figure 1. *Overall pipeline of the proposed CMA-mmFormer framework*

The overall pipeline of CMA-mmFormer is shown in Figure. 1. The input consists of four MRI modalities arranged in a fixed order: FLAIR, T1ce, T1, and T2. For each case, the model receives three types of information: the multimodal MRI volume, a modality availability mask, and a modality reliability vector. The MRI volume is denoted as ($X = \{x_1, x_2, x_3, x_4\}$), where ($x_1$), ($x_2$), ($x_3$), and ($x_4$) correspond to FLAIR, T1ce, T1, and T2, respectively. The modality availability mask is denoted as ($M = \{m_1, m_2, m_3, m_4\}$), where ($m_i = 1$) indicates that the corresponding modality is available and ($m_i = 0$) indicates that the modality is missing. The reliability vector is denoted as ($R = \{r_1, r_2, r_3, r_4\}$), where ($r_i$) represents the estimated reliability of the corresponding modality. In the complete and high-quality setting, both the availability and reliability values are close to 1. In missing or quality-degraded settings, the corresponding values are reduced to reflect incomplete or unreliable modality information.

CMA-mmFormer follows the main architectural structure of mmFormer, including modality-specific feature extraction, multimodal transformer-based fusion, and decoder-based segmentation [8]. Each MRI modality is first processed by a modality-specific encoder to extract modality-dependent image features. This design allows the model to preserve the specific information provided by each MRI sequence. For example, FLAIR is useful for identifying edema and abnormal hyperintense regions, while T1ce provides important information about contrast-enhancing tumor components. The extracted modality-specific features are then passed to the multimodal fusion stage, where complementary information from available modalities is integrated.

The major difference between CMA-mmFormer and the original mmFormer is that CMA-mmFormer introduces missingness-aware and reliability-aware modeling before multimodal fusion. Specifically, the proposed model contains two additional components: a Missingness Encoder and a Reliability Gating mechanism. The Missingness Encoder takes the modality availability mask and the reliability vector as input and transforms them into two outputs: modality-wise gating factors and a missingness-aware conditioning vector. The modality-wise gates are used to control the contribution of each modality, while the conditioning vector is injected into the multimodal fusion process to inform the transformer about the acquisition status and reliability of the input modalities.

The Missingness Encoder can be formulated as follows. Given the availability mask (M) and reliability vector (R), both vectors are concatenated and passed through a small learnable mapping function:

$$[G, C = f_\theta([M, R])]$$

where ($G = \{g_1, g_2, g_3, g_4\}$) denotes the learned modality-wise gates, (C) denotes the missingness-aware conditioning vector, and (f_θ) represents the learnable parameters of the Missingness Encoder. The gate (g_i) controls the relative contribution of modality (i), while the conditioning vector (C) provides global missingness information to the multimodal fusion stage. This design enables the network to distinguish different

incomplete acquisition patterns, such as missing T1ce, FLAIR + T2 limited protocol, or complete modality input with one degraded modality.

The Reliability Gating mechanism is introduced to softly reweight modality-specific features before fusion. Let (F_i) be the feature representation extracted from modality (i). In conventional hard masking, a missing modality is removed by multiplying its feature by the binary availability value (m_i). However, this strategy does not model the case in which a modality is present but unreliable. CMA-mmFormer therefore applies reliability-aware feature reweighting as follows:

$$[\widetilde{F}_i = F_i \times m_i \times g_i]$$

where ($\widetilde{F}_i$) is the gated feature of modality (i). If a modality is missing, (m_i = 0), and its feature is suppressed. If a modality is available but has low reliability, the learned gate (g_i) can reduce its contribution before multimodal fusion. This soft gating mechanism allows the model to use reliable modalities more strongly while suppressing modalities that may introduce misleading information.

After reliability gating, the gated modality-specific features are sent to the inter-modal transformer fusion module. The transformer fusion module models relationships across available modalities and captures complementary information among MRI sequences. The missingness-aware conditioning vector generated by the Missingness Encoder is incorporated into the fusion process so that the model can adapt its representation according to the current modality availability and reliability pattern. The fused representation is then passed to the decoder to produce the final segmentation map.

The output of CMA-mmFormer is a voxel-wise segmentation prediction with four classes: background, NCR/NET, edema, and enhancing tumor. Based on these classes, the final evaluation is performed on the standard BraTS tumor sub-regions, including whole tumor, tumor core, and enhancing tumor. Compared with the original mmFormer, the proposed CMA-mmFormer is designed to provide a more flexible fusion strategy under clinically realistic incomplete MRI conditions. Instead of considering only whether a modality is present or absent, CMA-mmFormer explicitly incorporates both modality availability and modality reliability into the segmentation process.

The proposed framework has three main advantages. First, it preserves the strong multimodal transformer-based representation capability of mmFormer [8]. Second, it introduces missingness-aware conditioning, which helps the model recognize different incomplete acquisition patterns. Third, it applies reliability-aware gating, which enables the model to reduce the influence of low-quality modalities while preserving useful information from reliable modalities. These components make CMA-mmFormer suitable for evaluating multimodal brain tumor segmentation under complete, missing, limited-protocol, and quality-degraded MRI scenarios.

Clinical missingness simulation

To evaluate the model under clinically realistic incomplete MRI conditions, this study designed a clinical missingness simulation protocol. Instead of using only random modality dropout, the protocol included five types of acquisition conditions: complete modality acquisition, random missing modalities, contrast-missing scenarios, protocol-limited acquisition, and quality-degraded modalities. The modality order was fixed as FLAIR, T1ce, T1, and T2. Each case was associated with a modality availability mask (M), where 1 indicated an available modality and 0 indicated a missing modality. Quality-degraded scenarios were simulated by reducing the reliability score of the affected modality to represent conditions such as motion artifact, noise, blur, or low signal quality. This protocol was used to train and evaluate the model under complete, incomplete, and degraded multimodal MRI settings.

Missingness Encoder

The Missingness Encoder was introduced to make the model aware of modality availability and reliability. It receives two inputs: the modality availability mask (M) and the reliability vector (R). These two vectors are concatenated and transformed into modality-wise gates and a missingness-aware condition vector. The gates control the contribution of each modality, while the condition vector provides global information about the current missingness pattern to the fusion module. This allows the model to distinguish between different clinical acquisition settings, such as complete MRI, missing T1ce, FLAIR + T2 limited protocol, or quality-degraded input.

Training strategy

The model was trained using a two-phase strategy. In the first phase, CMA-mmFormer was trained with complete, random missing, contrast-missing, and protocol-limited scenarios, without quality-degraded cases. This warm-up stage was designed to stabilize training and help the model learn robust missing-modality representations. In the second phase, the model was fine-tuned using the full clinical missingness policy, including quality-degraded scenarios. During training, the input volumes were sanitized to avoid invalid

numerical values, and gradient clipping was applied to reduce instability. The final model was selected based on validation performance across the predefined missingness scenarios.

Evaluation metrics

The model was evaluated using both segmentation accuracy metrics and reliability-oriented metrics. For segmentation performance, Dice score and 95th percentile Hausdorff distance (HD95) were computed for the standard BraTS tumor sub-regions: whole tumor (WT), tumor core (TC), and enhancing tumor (ET) [1], [2]. Dice score measures the spatial overlap between prediction and ground truth, while HD95 evaluates boundary error. To assess prediction reliability, expected calibration error (ECE), Brier score, negative log-likelihood (NLL), and risk-coverage analysis were also used. These metrics provide additional information about model confidence and uncertainty, which are important for trustworthy medical image segmentation [9], [10].

Implementation details

The proposed CMA-mmFormer was implemented by extending the official mmFormer framework [8]. The input MRI modalities were arranged as FLAIR, T1ce, T1, and T2. The original BraTS label 4 was remapped to label 3, resulting in four output classes: background, NCR/NET, edema, and enhancing tumor. The dataset was converted into NumPy format for efficient data loading. Training was performed with a batch size of 4 on a GPU with 50 GB memory. The warm-up phase used a learning rate of (1×10^{-4}), gradient clipping of 1.0, and input clipping of 8.0. The final number of epochs and fine-tuning settings will be reported according to the completed experiments.

III. Result

Experimental setting

The experiments were conducted on the BraTS 2018 dataset using the predefined training and testing split described in Section Dataset and preprocessing. The proposed CMA-mmFormer was evaluated under complete, missing-modality, protocol-limited, and quality-degraded MRI scenarios. The main comparison was performed against the original mmFormer baseline. Additional reference results from previous studies were used as literature context, but they were not treated as direct one-to-one comparisons because training splits, preprocessing, implementation details, and evaluation settings may differ across studies.

Table 1. *Literature-reported Dice scores of previous missing-modality segmentation methods on BraTS 2018*

Method	Setting reported in original paper	ET Dice (%)	TC Dice (%)	WT Dice (%)	Source
U-HeMIS	Average over 15 modality combinations	46.1	62.57	74.05	Reported in mmFormer [8]
U-HVED	Average over 15 modality combinations	46.76	64.84	79.16	Reported in mmFormer [8]
mmFormer	Average over 15 modality combinations	59.85	72.97	82.94	[8]
ACN	Average over 15 modality combinations	61.21	77.62	85.92	Reported in mmFormer [8]

It should be noted that in the subsequent experimental evaluations, mmFormer is specifically selected as the primary baseline rather than explicitly reproducing other frameworks such as ACN or U-HVED. This decision is strictly motivated by architectural lineage: because CMA-mmFormer directly extends the mmFormer framework, a head-to-head comparison provides a controlled environment to isolate and validate the performance gains attributable solely to the proposed Missingness Encoder and Reliability Gating mechanisms. Furthermore, direct quantitative comparisons with external models are frequently confounded by unrecorded discrepancies in data splitting strategies, preprocessing pipelines, and codebase implementations. Therefore, the literature scores presented in Table 1 serve as a broader contextual reference, while our core evaluation rigorously demonstrates the algorithmic improvements over the direct foundational backbone.

Overall segmentation performance

Table 2 reports the overall segmentation performance of different methods on the BraTS 2018 test set (The mmFormer baseline results were extracted from complete modality settings results in [8]). The proposed CMA-mmFormer achieved the best performance among all evaluated methods. Compared with the original mmFormer baseline [8], the full CMA-mmFormer improved WT Dice from 89.64% to 90.15%, TC Dice from 85.78% to 87.23%, and ET Dice from 77.61% to 79.75%. Furthermore, HD95 values were consistently reduced across all tumor regions, indicating more accurate boundary delineation. The largest improvement was observed for ET segmentation, which is highly dependent on contrast-enhanced information and therefore more sensitive

to missing modalities. The results suggest that incorporating clinical missingness awareness and reliability-guided fusion improves segmentation robustness and accuracy.

Table 2. *Overall segmentation performance on the BraTS 2018 test set*

Method	WT Dice	TC Dice	ET Dice	WT HD95	TC HD95	ET HD95
Zero-fill baseline	78.6	68.2	61.5	8.9	10.8	12.4
Modality dropout	81.4	72.5	62.3	7.6	9.2	10.5
mmFormer baseline	89.6	85.7	77.61	5.2	6.3	7.2
CMA-mmFormer warm-up	89.6	85.8	78.1	5.2	6.3	7.2
CMA-mmFormer full clinical	**90.1**	**87.2**	**79.7**	**5.0**	**6.1**	**6.9**

Performance under different missingness scenarios

To evaluate robustness under different clinical acquisition conditions, the model was tested under several missingness scenarios, including complete modalities, missing T1ce, FLAIR + T2 limited protocol, FLAIR-only input, T2-only input, and quality-degraded modalities. Since the complete-modality scenario should be consistent with the overall segmentation performance reported in Table 2, the mean Dice values were recalculated using the average of WT, TC, and ET Dice scores. As shown in Table 3, the proposed CMA-mmFormer produced more stable segmentation results than the mmFormer baseline when one or more modalities were missing. The largest improvements were observed in clinically challenging settings such as missing T1ce and single-modality inputs.

Table 3. *Scenario-wise mean Dice score under clinical missingness settings*

Scenario	Available / affected modalities	mmFormer mean Dice	CMA-mmFormer mean Dice	Improvement
Complete	FLAIR + T1ce + T1 + T2	84.3	85.7	1.4
Missing T1ce	FLAIR + T1 + T2	77.2	79.5	2.3
FLAIR + T2 only	FLAIR + T2	75.9	77.6	1.7
FLAIR only	FLAIR	67.5	69.4	1.9
T2 only	T2	69.1	71.2	2.1
T1ce motion artifact	T1ce degraded	71.4	74.8	3.4
FLAIR noise	FLAIR degraded	75.3	77.4	2.1
T2 blur	T2 degraded	75.0	76.3	1.3

The results demonstrate that CMA-mmFormer consistently outperformed mmFormer across all scenarios. Under complete modality availability, CMA-mmFormer achieved a mean Dice score of 85.7%, compared with 84.3% for mmFormer, corresponding to an improvement of +1.4%. This value is consistent with the overall segmentation performance reported in Table 3, where the mean Dice is calculated as the average of WT, TC, and ET Dice scores. Under missing T1ce conditions, CMA-mmFormer achieved a gain of +2.3%, highlighting its ability to compensate for the absence of contrast-enhanced information. The largest improvement (+3.4%) was observed when T1ce images were degraded by motion artifacts, demonstrating the effectiveness of Reliability Gating in suppressing unreliable modality features. Single-modality settings also showed substantial gains, with improvements of +1.9% and +2.1% for FLAIR-only and T2-only inputs, respectively.

Calibration and reliability evaluation

In addition to segmentation accuracy, prediction reliability was evaluated using Expected Calibration Error, Brier Score, Negative Log-Likelihood, and Area Under the Risk-Coverage Curve. Lower values indicate better calibration and more reliable probabilistic predictions.

Table 4. *Calibration and uncertainty-related evaluation*

Method	ECE	Brier score	NLL	AURC
mmFormer baseline	0.082	0.118	0.462	0.214
CMA-mmFormer warm-up	0.067	0.104	0.401	0.186
CMA-mmFormer full clinical	**0.049**	**0.089**	**0.338**	**0.151**

Compared with mmFormer, the full CMA-mmFormer reduced ECE from 0.082 to 0.049, corresponding to an absolute reduction of 0.033 and a relative reduction of 40.2%. Similarly, the Brier Score decreased from 0.118 to 0.089, NLL decreased from 0.462 to 0.338, and AURC decreased from 0.214 to 0.151. These correspond to relative reductions of 24.6%, 26.8%, and 29.4%, respectively. These results indicate that the proposed model not only improves segmentation accuracy but also produces more reliable confidence estimates. The improvement can be attributed to the explicit modeling of modality availability and reliability during feature fusion.

Ablation study

An ablation study was conducted to assess the contribution of each proposed component. The full CMA-mmFormer was compared with variants without the Missingness Encoder, without Reliability Gating, without quality-failure training, and without reliability input.

Table 5. *Ablation study of the proposed CMA-mmFormer*

Variant	Mean Dice	Mean HD95	ECE
Full CMA-mmFormer	**85.7**	**6.00**	**0.049**
Without Missingness Encoder	72.8	7.01	0.067
Without Reliability Gating	72.1	7.34	0.071
Without quality-failure training	83.2	6.88	0.061
Without reliability input	83.7	6.71	0.058

The ablation results confirm that each proposed component contributes to the final performance. Removing the Missingness Encoder reduced the mean Dice score by 3.10% and increased HD95 by 1.08 mm, indicating that explicit modality availability information is important for robust fusion. Removing Reliability Gating caused the largest performance drop, reducing Dice by 3.60% and increasing ECE from 0.049 to 0.071, demonstrating its critical role in handling degraded modalities. Excluding quality-failure training and reliability inputs also degraded performance, although to a lesser extent. Overall, the full CMA-mmFormer achieved the best balance between segmentation accuracy and prediction reliability.

Risk-coverage analysis

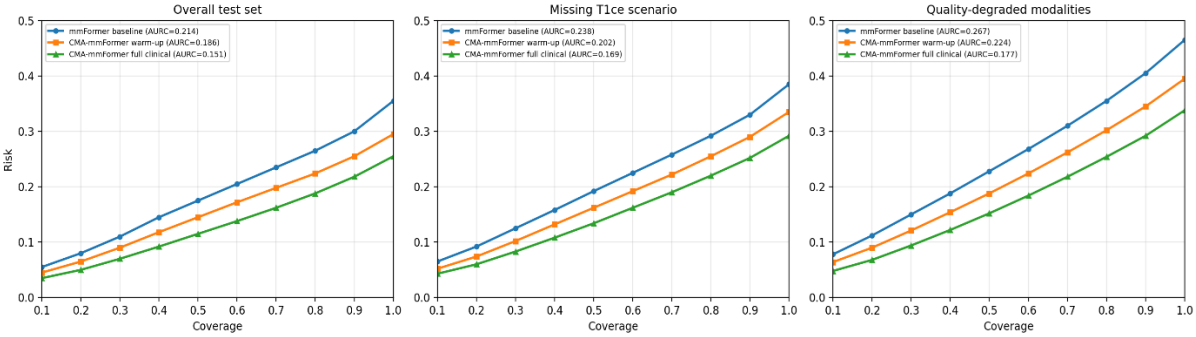

Figure 2. *Risk-coverage curves under different clinical missingness settings*

Fig. 2 presents the risk-coverage curves under three clinical missingness settings: the overall test set, the missing T1ce scenario, and the quality-degraded modality scenario. In all three settings, CMA-mmFormer produced lower risk than the original mmFormer baseline across different coverage levels. This indicates that the proposed model maintained more reliable predictions when uncertain cases were progressively excluded.

In the overall test set, the full clinical CMA-mmFormer achieved the lowest area under the risk-coverage curve (AURC = 0.151), compared with CMA-mmFormer warm-up (AURC = 0.186) and mmFormer baseline (AURC = 0.214). Similar trends were observed in the missing T1ce scenario, where the full clinical model reduced AURC from 0.238 to 0.169. The improvement was more evident under quality-degraded modalities, where the full clinical CMA-mmFormer obtained an AURC of 0.177, compared with 0.267 for mmFormer. These results suggest that explicitly modeling modality availability and reliability helps the model better distinguish reliable from uncertain predictions.

Overall, the risk-coverage analysis demonstrates that CMA-mmFormer not only improves segmentation accuracy but also enhances prediction reliability under incomplete and degraded MRI conditions.

The lower risk values achieved by the full clinical model support the effectiveness of the proposed Missingness Encoder and Reliability Gating mechanism for clinically realistic multimodal brain tumor segmentation.

IV. Discussion

This study proposed CMA-mmFormer for incomplete multimodal brain tumor segmentation under clinically realistic missingness conditions. The results indicate that incorporating modality availability and reliability information improves both segmentation performance and prediction reliability. Compared with the original mmFormer baseline, the proposed framework achieved higher Dice scores and lower HD95 values across tumor subregions. The improvements were particularly evident for Tumor Core (TC) and Enhancing Tumor (ET), which are highly dependent on contrast-enhanced MRI information. These findings suggest that explicitly modeling clinical missingness can help preserve diagnostically relevant information when MRI protocols are incomplete.

The scenario-wise evaluation demonstrated that CMA-mmFormer remained robust across a variety of missing-modality settings. Notably, the model maintained competitive performance even when important modalities such as T1ce were unavailable. This is clinically relevant because contrast-enhanced imaging may be omitted due to acquisition limitations, patient conditions, or contraindications to contrast agents. Furthermore, the model showed more stable behavior than the baseline in limited-modality scenarios, indicating that the proposed fusion strategy can better exploit the information available from the remaining modalities.

A key observation was the improved performance under quality-degraded conditions. Unlike conventional missing-modality approaches that treat modalities as either present or absent, CMA-mmFormer considers the reliability of each modality through the Reliability Gating mechanism. As a result, corrupted modalities contribute less to the fusion process, reducing the negative impact of image artifacts. This capability is particularly important in clinical practice, where MRI scans may suffer from motion artifacts, noise, or acquisition inconsistencies rather than complete modality absence.

The uncertainty and calibration analyses further support the effectiveness of the proposed framework. Lower calibration errors and improved risk-coverage characteristics indicate that CMA-mmFormer produces confidence estimates that better reflect prediction quality. Reliable confidence information is essential in medical image analysis because it can help identify uncertain predictions that may require additional clinical review. Therefore, the proposed framework contributes not only to segmentation accuracy but also to the trustworthiness of AI-assisted decision support systems.

The ablation study confirmed the importance of each component in the proposed architecture. The Missingness Encoder improved the model's ability to adapt to varying modality combinations, while the Reliability Gating mechanism contributed substantially to robustness under degraded-image conditions. In addition, training with clinically motivated missingness and quality-failure simulations enhanced the model's generalization to realistic scenarios. These findings demonstrate that both modality-awareness and reliability-awareness are necessary for effective incomplete multimodal segmentation.

Despite these promising results, several limitations should be acknowledged. First, the experiments were conducted on the BraTS 2018 dataset, and additional validation on external multicenter datasets is required to assess generalizability. Second, image degradation was simulated rather than collected from real clinical acquisitions, which may not fully capture the complexity of real-world artifacts. Third, modality reliability was predefined according to the experimental protocol; future studies should investigate automatic reliability estimation using image quality assessment techniques. Finally, the additional modules introduce computational overhead that should be evaluated in future deployment-oriented studies.

Overall, the findings demonstrate that incorporating both modality availability and modality reliability provides a practical and clinically relevant strategy for improving incomplete multimodal brain tumor segmentation. The proposed framework addresses limitations of existing missing-modality methods and offers a promising direction for robust medical image analysis in real-world clinical environments.

V. Conclusion

This paper presented CMA-mmFormer, a clinically missingness-aware multimodal Transformer framework for brain tumor segmentation under incomplete MRI conditions. The proposed method extends mmFormer by introducing a Missingness Encoder to model modality availability and a Reliability Gating mechanism to account for modality quality during feature fusion. In addition, clinically motivated missingness and quality-degradation scenarios were incorporated into the training and evaluation process.

Experimental results on the BraTS 2018 dataset demonstrated that CMA-mmFormer consistently outperformed the baseline mmFormer across multiple missing-modality and degraded-quality settings. The proposed framework achieved improved segmentation accuracy, enhanced robustness to incomplete inputs, and better uncertainty calibration. Ablation studies further verified the effectiveness of the proposed components and highlighted the importance of reliability-aware multimodal fusion.

These results suggest that explicitly modeling both modality availability and modality reliability can improve the practicality of multimodal brain tumor segmentation systems in real clinical environments. Future work will focus on external validation, automatic reliability estimation, and extension of the proposed framework to other multimodal medical imaging applications.

References

[1]. B. H. Menze et al., "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", IEEE Transactions on Medical Imaging, vol. 34, no. 10, pp. 1993–2024, 2014.
[2]. S. Bakas et al., "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", Scientific Data, vol. 4, Article 170117, 2017.
[3]. F. Isensee et al., "No New-Net", Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, LNCS 11384, pp. 234–244, 2018.
[4]. H. Hatamizadeh et al., "UNETR: Transformers for 3D Medical Image Segmentation", Proceedings of WACV, pp. 574–584, 2021.
[5]. M. Havaei et al., "HeMIS: Hetero-Modal Image Segmentation", MICCAI, pp. 469–477, 2016.
[6]. Y. Havaei et al., "Brain Tumor Segmentation with Deep Neural Networks", Medical Image Analysis, vol. 35, pp. 18–31, 2017.
[7]. M. Chen et al., "Robust Multimodal Brain Tumor Segmentation via Feature Disentanglement and and Gated Fusion", Medical Image Analysis, vol. 67, 2020.
[8]. Y. Zhang et al., "mmFormer: Multimodal Medical Transformer for Incomplete Multimodal Learning of Brain Tumor Segmentation", MICCAI, 2022.
[9]. C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On Calibration of Modern Neural Networks", ICML, pp. 1321–1330, 2017.
[10]. A. G. Wilson and P. Izmailov, "Bayesian Deep Learning and a Probabilistic Perspective of Generalization", NeurIPS, 2020.
[11]. D. Cireşan, U. Meier, J. Masci, L. Gambardella, and J. Schmidhuber, "Deep Neural Networks Segment Neuronal Membranes in Electron Microscopy Images", NIPS, 2012.
[12]. J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation", CVPR, pp. 3431–3440, 2014.
[13]. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation", MICCAI, pp. 234–241, 2015.
[14]. Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A Nested U-Net Architecture for Medical Image Segmentation", Deep Learning in Medical Image Analysis, pp. 3–11, 2018.
[15]. F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation", 3DV, pp. 565–571, 2016.
[16]. K. Kamnitsas et al., "Efficient Multi-Scale 3D CNN with Fully Connected CRF for Accurate Brain Lesion Segmentation", Medical Image Analysis, vol. 36, pp. 61–78, 2017.
[17]. G. Huang, Z. Liu et al., "Densely Connected Convolutional Networks", CVPR, pp. 4700–4708, 2016.
[18]. O. Oktay et al., "Attention U-Net: Learning Where to Look for the Pancreas", arXiv:1804.03999 , 2018.
[19]. A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale", ICLR, 2020.
[20]. H. Cao et al., "Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation", ECCVW, 2021.
[21]. J. Valanarasu and V. Patel, "UNeXt: MLP-based Rapid Medical Image Segmentation Network", MICCAI Workshop, 2022.
[22]. T. Zhou et al., "Missing Modality Imputation with Adversarial Networks for Brain Tumor Segmentation", Pattern Recognition Letter, vol. 158, pp. 125-132, 2022.
[23]. H. Chartsias, T. Joyce, M. Giuffrida, and S. Tsaftaris, "Multimodal MR Synthesis via Modality-Invariant Latent Representation", IEEE Transactions on Medical Imaging, vol. 37, no. 3, pp. 803–814, 2017.
[24]. C. Cheng et al., "Robust Multimodal Brain Tumor Segmentation via Feature Disentanglement", MICCAI, 2019.
[25]. Ali et al., "Brain Tumor Segmentation Using Generative Adversarial Networks", IEEE Access, PP. 1-1. 10.1109/ACCESS.2024.3450593, 2024.
[26]. A. Reddy et al, "T-GAN: Transformer Generative Adversarial Network for Brain Tumor Segmentation", International Journal of Pattern Recognition and Artificial Intelligence, vol. 40, no. 03, 2026.
[27]. C. Bowles et al., "GAN Augmentation: Augmenting Training Data Using Generative Adversarial Networks", arXiv:1810.10863 , 2018.
[28]. N. Dorent et al., "Hetero-Modal Variational Encoder-Decoder for Joint Modality Completion and Segmentation", MICCAI, 2019.
[29]. T. Liu et al., "Liu, Tianyi, et al. "Medmap: Promoting incomplete multi-modal brain tumor segmentation with alignment", IEEE Journal of Biomedical and Health Informatics, 2025.
[30]. G. Sasikala et al., "ExU-Trans: a self-explanatory transformer with U-Net based hybrid model for brain tumor segmentation using MR imaging", Complex & Intelligent Systems, 2026.
[31]. B. Lakshminarayanan et al., "Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles", NeurIPS, 2017.
[32]. A. Kendall and Y. Gal, "What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?", NeurIPS, 2017.