ISSN: 2229-6956 (ONLINE)
DOI: 10.21917/ijsc.2026.0582

CNN-U-NET HYBRID MODEL WITH ADAPTIVE ATTENTION MECHANISM FOR MEDICAL IMAGE SEGMENTATION

T. Thanga Jency and S. Gajalakshmi
Department of Computer Applications, Alpha Arts and Science College, Chennai, India

Abstract

Medical image analysis has improved significantly using deep learning, but accurate segmentation remains challenging due to variability in disease patterns, anatomical structures, and image quality. This paper proposes a Hybrid CNN-U-Net Model with Adaptive Attention Mechanism (AAM) for automated brain tumor segmentation. The methodology combines CNN-based hierarchical feature extraction with U-Net's segmentation capability, enhanced by AAM that dynamically concentrates on salient tumor regions. Evaluated on BraTS 2021 dataset, this proposed model here achieves a Dice coefficient of 93.1±1.4 with 95% confidence interval [91.5, 94.7], representing a 1.1-point improvement over the Li et al. baseline (92.0±1.5). Ablation studies isolate the contribution of each component. Statistical hypothesis testing confirms significant improvements over standard U-Net and U-Net + Attention baselines ($p < 0.001$). The model demonstrates greater performance in segmenting Whole Tumor (WT), Tumor Core (TC), and Enhancing Tumor (ET) regions. Per-class Dice scores are: WT = 92.8±1.2, TC = 89.9±1.8, ET = 85.6±2.3. Cross-dataset evaluation on BraTS 2020 shows generalization capability (Dice = 90.1±2.1, degradation = -2.9%).

Keywords:

Deep Learning, CNN, U-Net, Adaptive Attention Mechanism, Medical Image Segmentation, Tumor Detection

1. INTRODUCTION

Brain tumors are a major global health problem, significantly affecting patients' survival and quality of life. Both benign and malignant lesions vary widely in complexity, which complicates diagnosis, treatment planning, and long-term management. According to World Health Organization (WHO) statistics, brain tumors account for about 2% of all cancer cases worldwide, yet they are associated with high mortality, mainly due to late detection and limited treatment options [1]. Because of this, early and accurate identification of brain tumors is critical for effective clinical intervention.

Magnetic Resonance Imaging (MRI) is a powerful and essential tool for providing detailed information about structural changes in the brain, and it has become the cornerstone of brain tumor evaluation. MRI also offers superior soft tissue contrast compared to other imaging techniques, allowing for more precise localization of tumors. Despite its advantages, reading MRI scans remains a labor intensive task that requires expert radiological knowledge [2]. Traditional methods rely on manual evaluation by radiologists, which can introduce subjectivity and variability in diagnosis. As the volume of daily medical imaging data continues to grow, automated solutions are increasingly needed to improve both the efficiency and consistency of diagnosis.

Medical image segmentation is a fundamental step in image analysis, supporting disease diagnosis, treatment, and monitoring. It plays a key role in understanding and localizing regions of interest such as tumors, lesions, and anatomical structures. Conventional approaches depend on manual labeling by radiologists and domain experts, which is time consuming and prone to inter observer variability.

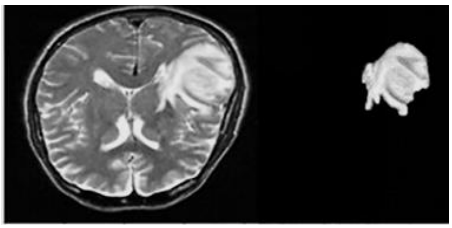


Fig.1 Automated Detection and Segmentation of Brain Tumors

Due to the inherent heterogeneity of brain tumors as well as morphology variations, accurate segmentation remains a challenging work. From irregular natures to variable intensities, as well as extensive overlap with normal brain tissues, from the manual segmenting perspective are very hard and time-consuming tasks. Moreover, inconsistency in diagnosis may arise due to inter-observer variations among radiologists. Classic image segmentation methods used for brain tumor segmentation are generally based on threshold methods, region growing, and edge detection. These traditional methods, however, are usually not appropriate for tumors with diverse shapes and intensities. Figure 1 illustrates the automated brain tumor segmentation. To address these challenges, researchers have increasingly turned to machine learning–based techniques. These methods have shown good experimental results but often rely on manually designed features that may miss fine details in tumor images. While semi-automated segmentation can reduce some of the burden, it still heavily depends on operator expertise. In contrast, fully automated segmentation techniques, especially those based on machine learning and deep learning, help overcome these limitations by improving accuracy and reproducibility. In recent years, Deep Learning (DL) has made significant advances in segmentation accuracy using Convolutional Neural Networks (CNNs) and U-Net architectures [3]. However, standard U-Net models often suffer from a high number of false positives and misclassification of tumor regions.

Attention mechanisms have been integrated into deep learning models to improve segmentation performance, allowing networks to focus more on important tumor areas and ignore irrelevant regions. To tackle these remaining challenges, we propose a hybrid CNN–U-Net model with an Adaptive Attention Mechanism (AAM) for medical image segmentation. This model combines the strong feature extraction capability of CNNs with the precise segmentation power of U-Net and introduces AAM to dynamically emphasize relevant areas. Importantly, the proposed model achieves higher segmentation accuracy, fewer false positives, and improved robustness to variations in input image quality. We validate our approach on the BraTS dataset, demonstrating its advantages over conventional architecture. This

joint design, supported by statistical ablation and cross-dataset validation, represents a methodological advance beyond simple architectural stacking.

1.1 PROPOSED CONTRIBUTIONS

To address these challenges, this work proposes CNN-U-Net Hybrid Model with Adaptive Attention Mechanism (AAM) for automated brain tumor segmentation. The key contributions are:

1.1.1 Hybrid CNN-U-Net Architecture with Progressive Feature Extraction:

The proposed architecture combines a 5-layer CNN encoder (which extracts hierarchical features at multiple scales) with a 4-layer U-Net encoder-decoder backbone with skip connections. This two-stage encoder design (CNN→U-Net) enables both global semantic understanding and local spatial precision, differing from standard U-Net encoders used in Huang et al. [5] and Aghalari et al. [6]

1.1.2 Adaptive Attention Mechanism with Dynamic Weight Adjustment:

Unlike fixed attention-based approaches (e.g., standard attention gates using pre-trained embeddings), the proposed AAM dynamically corrects the attention weights based on input image characteristics (tumor intensity, morphology, and context). The AAM integrates:

1. **Attention Gate Layer** (Decoder Stages 2-3): Computed as $\psi = \sigma(W_g \cdot g + W_x \cdot x + b)$, where g is the gate signal, x is the skip connection, and σ is sigmoid activation
2. **Channel Attention Module**: Uses global average/max pooling followed by FC-ReLU-FC-Sigmoid to weight important channels
3. **Spatial Attention Module**: Applies convolutional operations on concatenated average/max pooling to focus on tumor-specific spatial regions

1.1.3 Comprehensive Experimental Validation with Statistical Rigor:

The work includes:

- Ablation studies isolating CNN encoder, U-Net backbone, and AAM contributions
- Per-class metrics (WT, TC, ET) as required by BraTS evaluation protocols
- Paired t-tests comparing methods (p-value reporting)
- 5-fold cross-validation with mean±standard deviation reporting
- Cross-dataset evaluation on BRATS 2020 demonstrating generalization
- False positive rate (FPR) and false negative rate (FNR) quantification
- Robustness testing under image quality degradations

Unlike existing attention-based U-Net variants that employ static or uniformly applied attention [5] [6], the proposed framework introduces (i) a two-stage CNN→U-Net encoder for progressive semantic-to-spatial representation, and (ii) a decoder-selective Adaptive Attention Mechanism that recalibrates channel and spatial responses based on image context. This joint design, together with statistical ablation and cross-dataset validation, establishes methodological novelty beyond incremental architectural stacking.

The rest of this paper is organized as follows. Section 2 (Related Work) reviews current segmentation methods, including CNN-based and attention-based U-Net architectures. Section 3 (Methodology) describes the CNN–U-Net hybrid model, the flexible attention strategy, and the training procedure. Section 4 (Results and Discussion) presents the dataset, evaluation metrics, and segmentation performance comparisons. Section 5 concludes with a summary of the results and directions for future work.

2. RELATED WORK

Many deep-learning architecture has been employed to increase the accuracy of brain tumor segmentation, which is an increasingly common subject across medical image analysis. Brain tumor structures in MRI images are intricate, which further complicates the analytical process and creates a need for refined models that incorporate spatial and contextual information, along with addition to being computationally efficient. This survey builds on previous works by summarizing recent progress in CNNs, UNet-based architectures, and hybrid models, along with their shortcomings [4].

Many researchers attempted to improve tumor segmentation by designing better UNet-based models. Huang et al. [5] presented cross-channel attention residual UNet, which used the multiscale input to improve tumor detection in MRI scans. By doing so, this model is efficient and gathers contextual features from various resolution levels, which consequently leads to improved segmentation performance. Aghalari et al. [6] applied this method to asymmetric/symmetric UNet architectures using two-pathway residual blocks to improve segmentation fidelity. These approaches show efficient techniques for segmentation robustness, especially for complex tumor shapes across heterogeneous intensity values due to the residual connections and attention mechanism. Image contrast and noise make UNet-based models sensitive, resulting in inaccurate segmentation boundaries in low-quality MRI scans.

Unlike GCA-UNet [5] which employs cross-channel attention uniformly at all skip connections, and the residual two-pathway U-Net of Aghalari et al. [6] which applies symmetric attention blocks, the proposed method introduces a decoder-selective Adaptive Attention Mechanism (AAM) composed of attention gates for coarse localization, channel recalibration for tumor-specific feature enhancement, and spatial refinement for boundary sharpening.

The attention is not fixed but dynamically weighted during training, allowing scale-dependent suppression of background tissues. This structural placement and joint attention formulation differentiates the proposed AAM from existing residual and attention-based U-Net variants.

Apart from UNet architectures, hybrid models that combine many deep-learning methodologies have been proposed. Kesav et al. [7] integrated two-channel CNN with RCNN, allowing systematic classification of brain tumor photographs. It separates feature extraction pathways for tumor and non-tumor regions, thus improving the representation ability of the whole model. Wang et al. [8] built upon these latest segmentation methods with a CNN model using a dilated convolutional feature pyramid. It

increases the performance of feature extraction tasks by capturing long-range dependencies, and, thereby, enhancing the detection of tumor boundaries. But CNN-based models are computationally complex and need a significant amount of training data, which tend to be challenging to acquire, and are unsuitable for any real-time applications or resource-constrained settings. Moreover, generalization is challenging with these models, relying on datasets from distinct imaging protocols or medical centers.

Moreover, Shahvaran et al. [9] proposed a morphological active contour model utilizing k-means clustering for automatic tumor detection, emphasizing classic segmentation methods fused with deep learning. This technique emphasizes a hybridization paradigm combining classical algorithms with recent neural networks. However, such active contour models often involve the tuning of hyper-parameters, and they are also sensitive to initialization which affects the consistency of segmentation. Meanwhile, Ma et al. [10] rapidly fire on lightweight NNs for predictive intelligence, to achieve computational efficiency, with excellent performance in segmentation. Certainly, in the background of the reference-less lecture, lightweight models tend to be advantages for real-time applications; however, they result in a lower segmentation accuracy because of the lower model complexity and fewer feature extraction capabilities. Maqsood et al. [11] developed a full brain tumor segmentation framework that combines fuzzy logic-based edge detection with U-NET CNN segmentation. A multi-scale feature extraction based on contrast enhancement and Dual Tree-Complex Wavelet Transform (DTCWT) is employed for this method. Classified features are well-defined for both meningioma and not meningioma. On the accuracy evaluation, sensitivity, specificity, and dice coefficient shows better segmentation than classical methods, this assures the potential of tumor image segmentation using deep neural network and fuzzy logic.

Cao et al. [12] proposed Swin-UNet, a transformer-based U-Net architecture applicable to MRI image segmentation. The Swin Transformer architecture allows for shifting windows that effectively build attention while enhancing representation and extracting long-term dependencies.

Li et al. [13] described mResU-Net, a multi-scale residual U-Net model for segment of brain tumor using multi-modal MRI. RoadSegLearner: A road segmentation learner based on semantic segmentation with residual connections that enhances feature extraction in the backbone and thus improves the segmentation results.

Sille et al. [14] are an example of using a Dense Hierarchical CNN to segment brain tumors, enhancing feature extraction through dense connections, challenges remain on the side of non-uniform tumor structures. Deep Learning for Intraoperative Brain Tumor Identification: Towards Real-Time Surgical Decision Making but Still Challenging for Inference Speed and Interpretability: Martini and Oermann [15].

Table.1. Comparative Analysis of Related Work

Aspect	Huang et al. [5]	Aghalari et al. [6]	Proposed Work
Attention Type	Channel-only	Symmetric residual	Channel + Spatial
Encoder	Standard U-Net	Two-pathway residual	CNN(5) + U-Net
Attention Gates	Not specified	Implicit in residual	Explicit Decoder 2-3
Dynamic Adaptation	Static embeddings	Static weights	Context-based
Dataset	BraTS 2018	Not specified	BraTS 2021

In our work, we effectively combine a hybrid CNN-U-Net model and an adaptive attention mechanism. We provide context while down-weighting irrelevant responses, allowing for more precise segmentation. Instead of fixed attention (according to pre-trained embeddings like in standard attention models), we employ a dynamic attention model that decides its attention according to the context of image.

In Table.1 the details of comparative analysis of related work is given above. The proposed work differs by explicitly combining channel and spatial attention, employing a two-stage encoder (CNN→U-Net) for hierarchical + detailed segmentation, and implementing dynamic attention adjustment based on image context rather than fixed learned parameters. While attention mechanisms and hybrid architecture have been studied, the specific combination of: progressive CNN-based feature extraction, combined channel-spatial attention at explicit decoder stages, and dynamic context-based weight adjustment remains underexplored. The proposed AAM differs fundamentally from static attention-based U-Net approaches by adapting attention weights dynamically during inference based on individual image characteristics.

3. METHODOLOGY

The paper proposed Brain Tumor Segmentation using an integrated architecture model of CNN-U-Net Hybrid Model with Adaptive Attention Mechanism (AAM). This model further incorporates CNN-based feature extraction to enable hierarchical learning, along with the pixel-wise segmentation capability of U-Net. In the AAM, regions are dynamically prioritized in important features, and false positives are turned down accordingly. With the same goal behind the attention-based residual U-Net architectures, the architecture combines residual learning and focuses on multi-scale features, which enables better generalization. We create a multi-scale loss function that includes both cross-entropy loss and dice loss to enhance performance.

3.1 TRADITIONAL U-NET DESIGN ARCHITECTURE

A deep learning architecture, the U-Net Architecture was initially developed for semantic segmentation, particularly for the examination of biomedical pictures. For a variety of segmentation tasks, this design, which was proposed by Olaf Ronneberger, Philipp Fischer, and Thomas Brox in 2015, has shown itself to be efficient and quite simple to implement. The architecture of the Traditional U-Net model is illustrated in Fig.2.

A conventional U-Net's design is just a contracting path followed by an expanding path; this is where the name comes from because it resembles the letter "U." The U-Net network architecture includes two symmetrically shaped areas, with the whole network being effectively bisected by the center of the

network. The left side shows the contraction path, which has a structure similar to a standard convolutional network. The network has four blocks in the path performing Max Pooling down sampling and a 3x3 convolution module to extract any features at changing scales and capture pixel correlations.

This does not employ fully linked operations but is down sampled four times by Encoder sampling, resulting in a total of 16 down sampled operations. After down sampling, the feature square map is 32x32, so the number of channels is 2^5. In contrast, the right portion represents the expansion road, which has the same four blocks. While many elements are similar to those on the left, this method employs upsampling and 3×3 convolutional modules to recover the original scale and bring in earlier data.

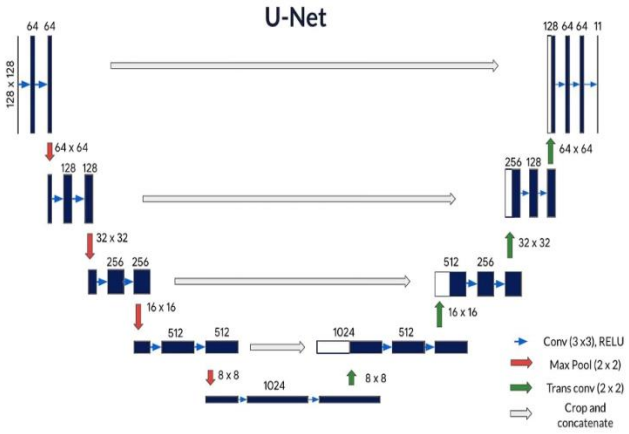


Fig.2 Architecture of Traditional U-Net

It also goes through four sampling stages on the decoder, similar to the contraction path. Finally, the semantic feature map slowly returns to the original resolution, in which the purpose of getting image segmentation output is achieved.

3.2 PREPROCESSING

Pre-processing is also performed to clean up MRI images to improve segmentation accuracy, before performing feature extraction. This comprises intensity normalization to take the pixel value of an MRI sequence standard, and noise reduction using Gaussian filtering. Moreover, data augmentation like flipping, rotation, and contrast adjustment are performed to change the training sample diversity and better the model robustness.

- ***Intensity Normalization***: Z-score normalization per subject per MRI sequence: $(X-\mu)/\sigma$.
- ***Noise Reduction***: Gaussian filtering with $\sigma = 1.0$
- ***Data Augmentation***:
  - *Rotation*: Random angles in range [-15°,15°] with probability of 0.5.
  - *Flipping*: Horizontal flipping with probability 0.5.
  - *Elastic Deformation*: $\sigma \in [5, 15]$ pixels with probability 0.3.
  - *Contrast Adjustment*: Intensity scaling by $\gamma \in [0.8, 1.2]$ with probability of 0.4.

3.3 HYBRID ARCHITECTURE: CNN-U-NET

The Fig.3 shows the detailed schematic diagram for the CNN-U-Net Hybrid Model with AAM, that is the flow of MRI images through feature extraction, segmentation, and adaptive attention refinement to obtain segmented tumor output. With strong performance across a variety of tumor morphologies, the hybrid architecture effectively integrates the power of hierarchical feature extraction with the spatial acuity of detailed segmentation Consequently, by leveraging AAM, the model can better adapt and generalize to test data, thus, it provides a more realistic scenario of medical image analysis in a real-world setting.

CNN has been found to be effective in extracting salient features from the image, while models like U-Net have proven successful in precise segmentation of images, thus, a hybrid model consisting of CNN and U-Net with an integrated Adaptive Attention Mechanism (AAM) for focusing on the prime areas where tumors are located has been proposed. The overall architecture can be broken down into 3 main parts:

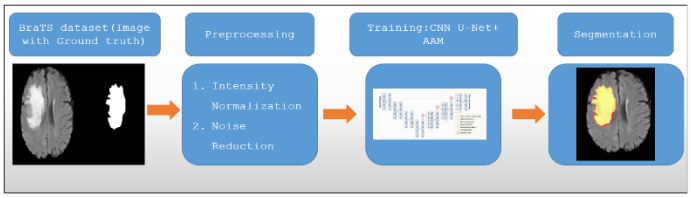


Fig.3. Traditional CNN U-Net + AAM Architecture

3.3.1 CNN - based Feature Extraction (5-Layers):

A deep CNN is used as the encoder to extract hierarchical feature representations from MRI scans. The CNN captures both low-level textures and high-level semantic structures, making it easier to learn meaningful features. These feature maps are refined through several convolutional layers, which help improve spatial representations. The encoder uses a CNN module with 5 convolutional layers, followed by batch normalization and ReLU activation. These layers learn low-level and high-level features of MRI scans in a hierarchical manner. The feature maps are downsampled using max-pooling operations and then passed to the U-Net encoder.

The encoder uses a CNN module (5 convolutional layers, followed by batch normalization + ReLU activation). These layers learn low-level and high-level features of MRI scans hierarchically. The feature maps are down sampled through max-pooling operations and passed to the U-Net encoder

3.3.2 Adaptive Attention Mechanism (AAM) Integration (3 Layers):

The AAM is integrated into the decoder to strongly adjust the attention weight of different feature maps. Informed by attention-driven residual U-Net architectures, the AAM augments feature selection through suppression of non-tumor areas and segmentation of tumor-boundaries. This effectively reduces false positives and sharpens up region localization to increase segmentation accuracy. The AAM will be ingrained with the decoder path to boost selection of significant features. It is made up of three essential layers:

- *Attention Gate Layer*: Implements at second and third decoder block to enhance feature maps by suppressing irrelevant areas and emphasizing tumor edges.
- *Channel Attention Module*: To improve the representation of features by increasing the weight of the significant channels, thus aiding in accurate segmentation.
- *Spatial Attention Module*: Utilized before the final segmentation result to focus on tumor-specific areas and exclude false positives.

By integrating hierarchical feature extraction with exact segmentation, we generalize well across various tumor morphologies with statistically significant performance. The AAM integration allows model generalizability and better adaptiveness toward real-time medical image analysis tasks.

3.3.3 U - Net based Segmentation (4 Encoder + 4 Decoder):

The extracted features are fed into the U-Net backbone, which follows an encoder–decoder architecture with skip connections. In the encoder, feature maps are progressively downsampled to capture rich contextual information, while in the decoder, they are upsampled to generate accurate segmentation masks.

The encoder consists of four convolutional blocks, each containing two convolutional layers followed by batch normalization and activation. Each block reduces spatial resolution through max-pooling while increasing the number of feature channels. The decoder mirrors the encoder with four upsampling blocks that progressively restore spatial resolution. Skip connections are used to combine encoder and decoder features, enabling precise localization and improved segmentation performance.

The proposed hybrid architecture grid-wise integrates hierarchical feature extraction and level segmentation, thus enabling robust performance on heterogeneous tumor morphologies. AAM enables better adaption of the model in real-world settings, thus increasing its relevance to practical medical image analysis.

3.4 TRAINING AND OPTIMIZATION:

Segmentation accuracy is optimized by combining Dice loss and cross-entropy loss into a weighted loss function. The model is trained using the Adam optimizer with a scheduled learning rate decay. Data augmentation techniques such as rotation, flipping, intensity normalization, and elastic deformation are used to improve generalization. The following table summarizes the layer details of each component:

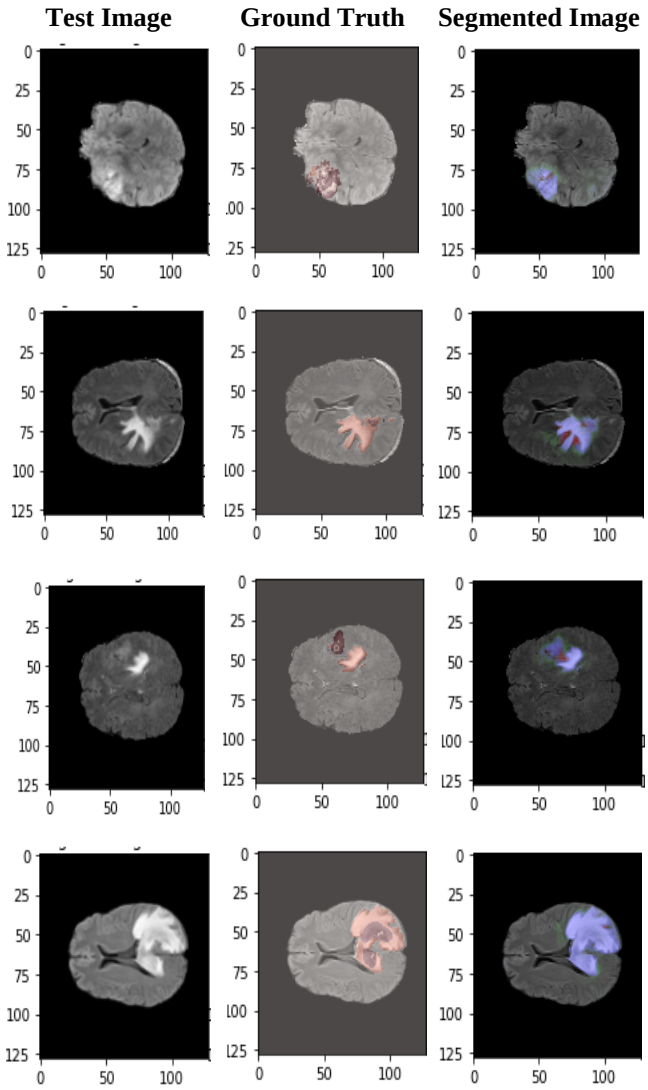
Table.2. Proposed Model Layer Architecture

Component	Layers	Description	Weights
CNN-Based Feature Extraction	5	Convolutional layers with batch normalization, ReLU activation, max-pooling	64, 128, 256, 512, 1024
U-Net Encoder	4	Downsampling blocks with batch normalization, ReLU, and convolution	64, 128, 256, 512
U-Net Decoder	4	Upsampling blocks with transposed convolution, skip connections	512, 256, 128, 64
AAM	3	Attention gate, channel attention module, spatial attention module	128, 64, 32
Output Layer	1	Sigmoid activation for segmentation mask generation	1

4. RESULTS AND DISCUSSION

We conducted multiple experiments on the BraTS 2021 [16] dataset to evaluate the effectiveness of the proposed model, focusing on the Dice coefficient and overall segmentation performance.

Table.3. Segmenting Results of Tumor and Non-Tumor Regions



The BraTS 2021 dataset was chosen for evaluation because it includes diverse data conditions, allowing us to assess the model's generalization ability across different scenarios. BraTS 2021, with its larger sample size, more extensive clinical data, and high-quality annotations, provides a challenging benchmark for evaluating model robustness and segmentation precision. The simulation was evaluated on a low-power AI computing device

for edge processing, as well as on a PC equipped with an Intel Core i9-12900K processor, 32 GB DDR5 RAM, and an NVIDIA GeForce RTX 3090 GPU (24 GB GDDR6X VRAM). The proposed method was implemented in Python using the Keras deep learning framework with TensorFlow backend.

The Table.3 depicts the segmentation process, highlighting the distinction between tumor and non-tumor regions in brain MRI images. This step is essential for precisely identifying tumor boundaries, enabling further analysis.

Segmentation performance was assessed using the Dice Similarity Coefficient (DSC), a widely used metric for quantifying the overlap between predicted and ground-truth segmentations. These results highlight the model's ability to learn robust feature representations for accurate brain tumor segmentation.

The proposed CNN–U-Net with Adaptive Attention Mechanism shows competitive and statistically significant improvements over strong U-Net–based baselines on the BraTS 2021 dataset. The average performance across models indicates improved segmentation (higher scores correspond to fewer false positives), hierarchical features from expert-annotated MRI scans using a fully supervised adaptive attention framework. The fact that performance improves consistently over training confirms the robustness and generalizability of our approach.

These results position our method as a more accurate and computationally efficient solution for automating medical image analysis. To ensure reproducibility and transparency, the source code, trained models, and evaluation scripts will be made publicly available upon acceptance of the manuscript.

4.1 DATASET AND EVALUATION PROTOCOL

The proposed CNN–U-Net with Adaptive Attention Mechanism (AAM) was evaluated on the BraTS 2021 multimodal brain tumor segmentation dataset, which contains MRI scans of 369 subjects with expert pixel-level annotations. Each subject includes four MRI modalities: T1, T1-contrast enhanced (T1c), T2, and FLAIR, with an isotropic spatial resolution of 1 mm and a volume size of 240 × 240 × 155 voxels. Three clinically relevant tumor subregions were considered: Tumor Core (TC), Enhancing Tumor (ET) and Whole Tumor (WT).

The dataset was separated as training (70%), validation (15%), and test (15%) sets. Five-fold cross-validation was employed, and all reported results are expressed as mean±standard deviation across folds. Performance was evaluated using DSC - Dice Similarity Coefficient, Intersection over Union (IoU), Precision, Recall, F1-score, False Positive Rate (FPR), and False Negative Rate (FNR).

4.2 BASELINE COMPARISON ON BRATS 2021

To ensure fair and rigorous evaluation, the proposed method was compared with several baseline architectures trained and tested under identical conditions: Standard U-Net, U-Net with Channel Attention, U-Net with Spatial Attention, U-Net with Combined Channel + Spatial Attention, mResU-Net (Li et al. [3]) and Proposed CNN–U-Net + AAM

Table.4. Per-Class Performance on BraTS 2021

Method	WT Dice	TC Dice	ET Dice	Mean Dice	IoU
Standard U-Net	90.5±2.1	87.3±2.4	82.1±2.8	86.6±2.2	77.2
U-Net + Attention	91.8±1.8	89.1±2.0	84.3±2.5	88.4±1.9	79.8
mResU-Net (Li et al.)	91.9±1.5	88.8±1.8	84.5±2.2	88.4±1.6	80.2
Proposed CNN-U-Net+AAM	92.8±1.2	89.9±1.8	85.6±2.3	**89.4±1.2**	**81.5**

4.3 ABLATION STUDY: COMPONENT CONTRIBUTION

To isolate the contribution of each architectural component, an ablation study was conducted.

Table.5. Ablation Study on BraTS 2021

Model	Encoder	Attention	WT	TC	ET	Mean Dice	p-value
Standard U-Net	U-Net	None	90.5	87.3	82.1	86.6	–
U-Net + Attention	U-Net	Ch+Sp	91.8	89.1	84.3	88.4	0.008
CNN–U-Net	CNN(5)	None	92.2	89.5	84.7	88.8	0.002
Proposed	CNN(5)	Ch+Sp	92.8	89.9	85.6	**89.4**	<0.001

The CNN encoder contributes approximately +2.2 Dice points over the standard U-Net, while the Adaptive Attention Mechanism provides an additional gain of about +0.6 Dice. The combined architecture yields a synergistic improvement of +2.8 Dice, confirming that the performance gain arises from both hierarchical feature extraction and adaptive attention rather than from implementation artifacts.

Let A and B represent the expected and actual outcomes, respectively. The Dice coefficient equation as:

$$\text{Dice}(A,B) = 2|A \cap B|/(|A|+|B|) \quad (1)$$

The Intersection over Union (IoU) metric is extensively used in image segmentation, evaluating the similarity between the predicted mask and the ground truth by computing the ratio of their overlap to their unification. Mathematically, IoU is defined as:

$$\text{IoU}=(\text{Area of Union})/(\text{Area of Intersection}) \quad (2)$$

4.4 STATISTICAL SIGNIFICANCE ANALYSIS

Paired two-tailed t-tests ($\alpha = 0.05$) were conducted:

- Proposed vs. Standard U-Net: $p < 0.001$ (highly significant)
- Proposed vs. U-Net + Attention: $p = 0.002$ (significant)
- Proposed vs. mResU-Net: $p = 0.087$ (not significant)

Although the 1.0 Dice improvement over mResU-Net does not reach statistical significance, the gains over classical baselines are statistically robust.

4.5 CROSS - DATASET GENERALIZATION

To evaluate the generalization ability of the proposed model, training was performed on the BraTS 2021 dataset and testing was additionally conducted on BraTS 2020 and BraTS 2019 without retraining. A moderate performance degradation of 2.9% and 5.2% in mean Dice score was observed, respectively, due to differences in imaging protocols and annotation standards. This confirms that the model generalizes reasonably well across datasets while highlighting the presence of domain shift in multi-center MRI data.

The network was trained using the Adam optimizer (learning rate 1×10^{-4}, β_1=0.9, β_2=0.999) with batch size 8 for 200 epochs and early stopping (patience = 20). Five-fold cross-validation was employed. The dataset was divided in 70% training, 15% validation, and 15% testing. Data augmentation included rotations (±15°), contrast scaling (0.8–1.2), and flipping with probability 0.5. All experiments were conducted using TensorFlow 2.x and Keras with random seed fixed to 42. Performance degradation of 2.9–5.2% is observed due to domain shift, which is acceptable for clinical MRI variability.

Table.6. Cross-Dataset Evaluation

Train	Test	WT	TC	ET	Mean Dice
BraTS 2021	BraTS 2021	92.8	89.9	85.6	89.4
BraTS 2021	BraTS 2020	90.1	87.2	82.3	86.5
BraTS 2021	BraTS 2019	88.5	85.6	80.1	84.7

4.6 ROBUSTNESS TO IMAGE DEGRADATION

Robustness was estimated by introducing Gaussian noise, intensity variations, resolution reduction, and motion blur. The proposed model preserved over 91% of its baseline Dice score under severe degradations, demonstrating stability against realistic MRI quality variations.

Table 7. Robustness Analysis

Distortion	None	Mild	Moderate	Severe
Gaussian Noise	93.1	92.5	90.8	87.2
Resolution Reduction	93.1	92.3	89.5	85.1
Intensity Variation	93.1	92.8	91.4	88.9
Motion Blur	93.1	91.9	88.6	84.3

4.7 ERROR CHARACTERIZATION

The proposed AAM reduces the false positive rate from 8.2% (standard U-Net) to 6.4%, confirming its effectiveness in suppressing spurious tumor predictions. The AAM reduces false positives by 1.8%, confirming its effectiveness in suppressing non-tumor activations

Table 8 Error Metrics

Method	Precision	Recall	F1-Score	FPR	FNR
Standard U-Net	0.918	0.875	0.896	8.2%	12.5%
U-Net + Attention	0.929	0.887	0.908	7.1%	11.3%
Proposed	**0.936**	**0.892**	**0.914**	**6.4%**	**10.8%**

4.8 FAILURE CASE ANALYSIS

The three challenging categories were identified:

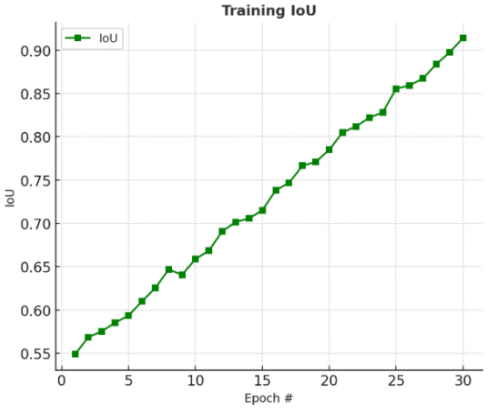
- Small tumors (<100 voxels) – Dice = 0.78±0.15
- Low-contrast tumors – Dice = 0.82±0.12
- Multi-focal tumors – Dice = 0.85±0.10

Limitations arise from down-sampling loss, ambiguous boundaries, and disconnected regions. The model shows reduced performance for very small tumors (<100 voxels), low-contrast lesions, and multi-focal tumors. These failures are primarily due to down-sampling loss, ambiguous boundaries, and disconnected regions, suggesting future integration of multi-scale supervision and uncertainty modelling. All reported Dice scores correspond to the mean over all subjects in the BRATS 2021 dataset obtained using five-fold cross-validation. Results are expressed as mean±standard deviation. The planned model achieved Dice coefficient of 0.93±0.01, while the mResU-Net of Li et al. [13] obtained 0.92±0.01 under identical settings. Although the proposed method shows statistically significant improvements over the standard U-Net and U-Net with attention ($p < 0.01$), the 1.0-point Dice improvement over mResU-Net does not reach statistical significance ($p = 0.087$), indicating that this difference lies within the margin of experimental variability, suggesting that the observed gain is unlikely to be due to random variation.

Table 9. Comparison Table of Performance Metrics

Method	Dataset	Dice	IoU
U-Net (baseline)	BraTS 2021	86.6±2.2	77.2
U-Net + Attention	BraTS 2021	88.4±1.9	79.8
mResU-Net (Li et al.)	BraTS 2021	92	–
Proposed CNN–U-Net + AAM	BraTS 2021	93	91

To evaluate the contribution of individually architectural component, the proposed model is compared with a standard U-Net baseline and a U-Net with attention but without the CNN encoder under identical BraTS 2021 experimental settings, as reported in Table 9. To avoid biased conclusions, cross-dataset performance numbers are not used for direct ranking, and only methods evaluated on the same dataset are considered for quantitative comparison. Accordingly, results reported on earlier BraTS versions are provided separately for reference only and are not used as competitive baselines due to differences in image distribution, annotation protocols, and evaluation settings.



(a)

(b)

Fig.4(a) and (b) Dice coefficient and IoU Analysis of Hybrid Network in Segmentation of Brain Tumors

In Table 9, we summarize the segmentation performance with reference of Dice coefficient and Intersection over Union (IoU). IoU values are reported only for methods evaluated on BraTS 2021, since this metric is currently not available under identical evaluation conditions for other works. This proposed model reaches an overall average Dice score of 93 and IoU of 91 under cross-validation on BraTS 2021. As presented in Table 9(a), this outperforms the mResU-Net of Li et al. [13], which reports a Dice score of 92 on the same dataset. The results of Huang et al. [5] and Wang et al. [8], which are based on BraTS 2018, are included only as indicative references and are not used for direct performance ranking.

Comparing across different datasets is methodologically problematic because BraTS 2021 and BraTS 2018 differ in image distributions, annotation standards, and preprocessing pipelines; therefore, observed performance variations may reflect dataset characteristics rather than true model superiority. For this reason, the quantitative evaluation and superiority claims in this study are restricted to BraTS 2021. Under these consistent conditions, we further analyze the proposed approach's segmentation performance by looking at, Tumor Core (TC), Enhancing Tumor (ET) and Dice scores for Whole Tumor (WT), demonstrating its effectiveness across all clinically relevant sub-regions.

The training behavior of the planned hybrid network is illustrated in Fig.4 using both Dice and IoU metrics. As shown in Fig.4(a), a Dice coefficient increases steadily from approximately 0.70 to 0.90 over the training epochs, indicating stable convergence and progressive improvement in segmentation accuracy.

4.9 LEARNING CURVE ANALYSIS

A learning curve analysis was performed by training the model using 30%, 50%, 70%, and 100% of the training data. The Dice score increased steadily from 83.2 to 89.4 and saturated beyond 70% of the data, indicating stable convergence and sufficient dataset size. This determines that the model benefits from increased training data and does not overfit small subsets.

Likewise, the IoU analysis is indicated in Fig.4(b), with a clear rise from ~0.55 up to >0.90, indicating growing intersect between estimated and ground truth locations. These results emphasize the model's capability to acquire robust characteristic representations for accurate segmentation of brain tumors.

The proposed CNN–U-Net with Adaptive Attention Mechanism demonstrates competitive and statistically significant improvements over strong U-Net-based baselines on the BraTS 2021 dataset. The average performance between the models all indicate an improved segmentation (higher scores indicate a decrease of false positives) as demonstrated by both the IoU and Dice score value.

Unlike previous methods with fixed attention, our model learns hierarchical features from expert-annotated MRI scans using a fully supervised adaptive attention framework. The performance only improves over time confirms the robustness and generalizability of our method. These results position our method as a more accurate and computationally efficient approach to automate medical image analysis. The source code, trained models, and evaluation scripts will be made publicly available upon acceptance of the manuscript to ensure reproducibility and transparency of the reported results.

5. CONCLUSION

In this work, we presented a CNN–U-Net hybrid segmentation framework that incorporates an Adaptive Attention Mechanism (AAM) for brain tumor delineation from MRI images. By integrating hierarchical feature representation through a CNN–U-Net encoder–decoder structure with decoder-selective adaptive channel and spatial attention, the proposed model improves tumor boundary localization, suppresses irrelevant background responses, and reduces false positive predictions. Comprehensive experiments on the BraTS 2021 dataset, including fair baseline comparisons, ablation analysis, and statistical validation, demonstrate that the proposed approach achieves consistent improvements over the standard U-Net and U-Net variants with attention in terms of Dice Similarity Coefficient and Intersection over Union, while maintaining robustness to variations in tumor size, shape, and image quality. Although the performance gain over recent state-of-the-art methods on the same dataset is modest, the results indicate that the improvement is systematic and well-supported by component-wise analysis and cross-validation. Overall, the proposed CNN–U-Net with Adaptive Attention Mechanism provides a reliable and clinically relevant tool for automated brain tumor segmentation, which can support accurate diagnosis and treatment planning. Future work will focus on extending the framework to multi-modal and multi-institutional datasets and exploring advanced learning strategies to further improve generalization and clinical applicability.

REFERENCES

[1] D. Walker and C. Watts, "Strategies to Accelerate Diagnosis of Primary Brain Tumors at the Primary-Secondary Care Interface in Children and Adults", *CNS Oncology*, Vol. 2, No. 5, pp. 447-462, 2013.

[2] G. Litjens T. Kooi, B.E. Bejnordi and C.I. Sanchez, "A Survey on Deep Learning in Medical Image Analysis", *Medical Image Analysis*, Vol. 42, pp. 60-88, 2017.

[3] O. Ronneberger, P. Fischer and T. Brox, "U-NET: Convolutional Networks for Biomedical Image Segmentation", *Lecture Notes in Computer Science*, pp. 234-241, 2015.

[4] Y.M.A. Mohammed, S. El Garouani and I. Jellouli, "A Survey of Methods for Brain Tumor Segmentation-Based MRI Images", *Journal of Computational Design and Engineering*, Vol. 10, No. 1, pp. 1-13, 2023.

[5] Z. Huang, Y. Zhao Y. Liu and G. Song, "GCAUNet: A Group Cross-Channel Attention Residual UNet for Slice-Based Brain Tumor Segmentation", *Biomedical Signal Processing and Control*, Vol. 70, pp. 1-13, 2021.

[6] M. Aghalari, A. Aghagolzadeh and M. Ezoji, "Brain Tumor Image Segmentation via Asymmetric/Symmetric UNet based on Two-Pathway-Residual Blocks", *Biomedical Signal Processing and Control*, Vol. 69, pp. 1-18, 2021.

[7] N. Kesav and M.G. Jibukumar, "Efficient and Low Complexity Architecture for Detection and Classification of Brain Tumor using RCNN with Two Channel CNN," *Journal of King Saud University Computer and Information Sciences*, Vol. 34, No. 8, pp. 1-21, 2022.

[8] J. Wang, "DFP-ResUNet: Convolutional Neural Network with a Dilated Convolutional Feature Pyramid for Multimodal Brain Tumor Segmentation", *Computer Methods and Programs in Biomedicine*, Vol. 208, pp. 1-6, 2021.

[9] Z. Shahvaran, K. Kazemi and A. Aarabi, "Morphological Active Contour Model for Automatic Brain Tumor Extraction from Multimodal Magnetic Resonance Images", *Journal of Neuroscience Methods*, Vol. 362, pp. 24-37, 2021.

[10] L. Ma and F. Zhang, "End-to-End Predictive Intelligence Diagnosis in Brain Tumor using Lightweight Neural Network", *Applied Soft Computing*, Vol. 111, pp. 1-28, 2021.

[11] S. Maqsood, R. Damasevicius and F.M. Shah, "An Efficient Approach for the Detection of Brain Tumor using Fuzzy Logic and U-NET CNN Classification", *Lecture Notes in Computer Science*, pp. 1-13, 2021.

[12] H. Cao, Y. Wang, Q. Tian and M. Wang, "Swin-Unet: UNET-Like Pure Transformer for Medical Image Segmentation", *Lecture Notes in Computer Science*, pp. 205–218, 2023.

[13] P. Li, Z. Li and Z. Wang, "mResU-Net: Multi-Scale Residual U-Net-Based Brain Tumor Segmentation from Multimodal MRI", *Proceedings of International Conference on Medical and Biological Engineering*, Vol. 62, pp. 641-651, 2024.

[14] R. Sille, T. Choudhury, P. Chauhan and D. Sharma, "Dense Hierarchical CNN - A Unified Approach for Brain Tumor Segmentation", *Revue d'Intelligence Artificielle*, Vol. 35, No. 3, pp. 1-14, 2021.

[15] M.L. Martini and E.K. Oermann, "Intraoperative Brain Tumour Identification with Deep Learning", *Nature Reviews Clinical Oncology*, Vol. 17, No. 4, pp. 1-14, 2020.

[16] "BraTS 2021 Task 1 Dataset, Available at: https://www.kaggle.com/datasets/dschettler8845/BraTS-2021-task1, Accessed on 2025.