Journal of Electronics, Electromedical Engineering, and Medical Informatics
Homepage: jeeemi.org; Vol. 8, No. 4, October 2026, pp: 1389-1405 e-ISSN: 2656-8632
RESEARCH ARTICLE OPEN ACCESS

Adaptive Sparse Cross-Scale Transformer for Computationally Efficient Brain Tumor MRI Segmentation

RaviChandra Bandi[1], Selvanayaki SundarRaj[2], Amitha I. C.[3], Sudha Subramaniam[4], Suganthi Raja[5], and Anand Rajendran[6]

[1] Department of ECE, N S Raju Institute of Technology (A), Sontyam, Visakhapatnam, India
[2] Department of Artificial Intelligence and Data Science, Saveetha Engineering College, Chennai, India
[3] Department of Computer Science and Engineering, KPR Institute of Engineering and Technology, Coimbatore, India
[4] Department of ECE, Kongu Engineering College, Perundurai, India
[5] Department of Electronics and Communication Engineering, Panimalar Engineering College, Chennai, India.
[6] Department of Aeronautical Engineering, Nehru Institute of Technology, Coimbatore, India

Corresponding author: RaviChandra Bandi (e-mail: ravichand678@gmail.com), **Author(s) Email**: Selvanayaki SundarRaj (e-mail: selvanayakis@saveetha.ac.in), Amitha Ida Chandran (e-mail: amitha.ic@kpriet.ac.in), Sudha Subramaniam (e-mail: sudha.ece@kongu.edu), R.Suganthi (e-mail: sugimanicks@gmail.com), Anand Rajendran (e-mail: anand.st2010@gmail.com)

Abstract **Brain tumor segmentation from Magnetic Resonance Imaging (MRI) is a critical task in medical image analysis, owing to the irregular morphology of tumors, heterogeneous intensity distributions across MRI modalities, and the presence of multiple overlapping sub-regions. Accurate delineation of these regions is essential for clinical diagnosis, treatment planning, and longitudinal disease monitoring. Although Vision Transformer (ViT)-based architectures have demonstrated promising performance in medical image segmentation by capturing long-range dependencies and global contextual information, conventional multi-scale transformer models remain constrained by static grouped attention mechanisms that introduce redundant computations and exhibit quadratic complexity with respect to input size. To address these limitations, this paper proposes the Adaptive Sparse Cross-Scale Transformer (ASCT), a computationally efficient framework for brain tumor MRI segmentation. ASCT incorporates three key innovations: (i) a Dynamic Scale Routing (DSR) module that adaptively weights multi-scale features using learned routing coefficients, replacing fixed channel grouping; (ii) a Sparse Token Attention (STA) mechanism that restricts attention computation to the most informative token pairs, reducing complexity from quadratic $O(N^2)$ to near-linear $O(Nk)$; and (iii) a linearized attention approximation that significantly reduces GPU memory consumption during training. Additionally, cross-scale feature fusion is performed prior to the attention operation to suppress redundant computations and enhance inter-scale feature interaction. The proposed ASCT model is evaluated on the BraTS 2021 dataset for multi-region segmentation, targeting Whole Tumor (WT), Tumor Core (TC), and Enhancing Tumor (ET) sub-regions. Experimental results demonstrate that ASCT achieves Dice scores of 89.2%, 84.1%, and 83.5% for WT, TC, and ET, respectively, yielding an average Dice score of 85.6%. Compared to the baseline Vision Transformer, the proposed model reduces computational complexity from 94.6 GFLOPs to 58.4 GFLOPs and GPU memory usage from 11.2 GB to 7.3 GB, confirming its efficiency and practical viability for real-world clinical applications.**

Keywords **Brain tumor segmentation, BraTS, Sparse attention, MRI segmentation, Transformer-based segmentation.**

I. Introduction

Recent transformer-based medical image segmentation models have significantly improved the ability to capture long-range contextual information compared with conventional CNN architectures. However, their practical deployment remains challenging due to substantial computational and memory requirements. For example, UNETR employs global self-attention over all image tokens, resulting in quadratic computational complexity with respect to token count. As image resolution increases, the attention computation and memory consumption

Manuscript received 9 May 2026; Revised 30 June 2026; Accepted 20 July 2026; Available online 13 August 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1742

increase rapidly, limiting scalability for high-resolution MRI segmentation [1]. Similarly, TransUNet combines CNN feature extraction with transformer encoders to improve contextual representation, but still relies on dense self-attention operations and multi-stage feature processing, leading to considerable computational overhead. Swin-UNet partially addresses this issue through shifted-window attention, reducing attention computation within local windows. Nevertheless, global contextual interactions require multiple hierarchical stages, and the model still performs attention independently at different feature scales, introducing redundant computations [2]. More recent cross-scale transformer architectures attempt to improve multi-resolution feature interaction through cross-attention or hierarchical fusion mechanisms. Although these approaches enhance feature representation, they typically compute attention separately at each scale and subsequently aggregate the results. Such designs increase computational redundancy, parameter count, and memory usage, particularly when processing large medical images containing thousands of tokens. Furthermore, most existing methods employ fixed channel grouping or predefined scale-interaction strategies, which cannot dynamically adapt to the varying sizes, shapes, and appearances of tumor regions across patients [3].

To address these limitations, the proposed Adaptive Sparse Cross-Scale Transformer (ASCT) introduces three complementary innovations. First, Dynamic Scale Routing (DSR) adaptively assigns importance weights to different feature scales, replacing static channel grouping and enabling data-driven scale selection [4]. Second, Sparse Token Attention (STA) restricts attention computation to the most informative token pairs, reducing complexity from $O(N^2)$ to approximately $O(Nk)$ while preserving critical contextual relationships. Third, Cross-Scale Pre-Attention Fusion combines multi-scale features before attention computation, eliminating the need for repeated attention operations across individual scales. In addition, a Linearized Attention Approximation further reduces memory consumption during training and inference. Unlike UNETR, TransUNet, Swin-UNet, and existing cross-scale transformers, ASCT simultaneously addresses computational efficiency, memory scalability, and adaptive multi-scale interaction within a unified segmentation framework [5]. The main contributions of this work are summarized as follows:

1. A dynamically changing scale routing system that adaptively weights features over multiple scales instead of fixed channel grouping.
2. A sparse token-based attention mechanism that reduces computational costs from quadratic to nearly linear.
3. An approximation of attention that linearizes the process to greatly reduce memory usage in training.
4. An efficient transformer-based framework for segmentation that has been tested with BraTS 2021 and gives improved Dice scores and reduced FLOPs and GPU memory.

The paper is organized in the following manner: Section 2 reviews previous studies that have employed CNN and transformers in segmenting brain tumors. Section 3 discusses the ASCT framework with its architecture. Section 4 describes the experiments conducted and presents quantitative results using the BraTS dataset. Section 5 is dedicated to a discussion of the results with an analysis of computational efficiency. Finally, Section 6 concludes the work and gives directions for future research.

II. State-of-the-Art Techniques

Deep learning-based classification and segmentation models have predominantly led to recent developments in the analysis of brain tumors. The classic convolutional neural network (CNN) models enhance discrimination using deeper customized architectures and more advanced feature hierarchies, but are still fundamentally limited by smaller receptive fields and locality bias, which limit the global contextual relationships in MRI scans [6]. Explainable CNN architectures are better in terms of interpretability since they emphasize salient parts. However, they do not fundamentally resolve the limited long-range modeling power of convolutional operations [7]. Hybrid CNN SVM methods that are optimized with metaheuristic algorithms like PSO are trying to improve the classification ranges by way of the feature classifier division, albeit the execution relies greatly on handcrafted optimization and omits the explicit global interaction modelling [8]. ESimilarly, other optimized hybrid deep learning models typically focus on incremental accuracy improvements through tuning strategies rather thanstructural efficiency or scalable contextual learning [9], [10].

To address the locality constraint, a number of recent studies adopt a combination of transformers and CNN backbones in order to learn global relationships. End-to-end CNN - Vision Transformer models enhance semantic representation capability by fusing local and global features, but in many cases, the hybrids are more expensive to execute, and fail to explicitly address redundant multi-scale processing [11]. Models with CNN--Transformer synergy and pre-trained transformer-based strategies improve contextual awareness and are able to take advantage of the benefits of transfer learning, but still have quadratic complexity due to self-attention and are computationally expensive for high-resolution MRI

Manuscript received 9 May 2026; Revised 30 June 2026; Accepted 20 July 2026; Available online 13 August 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1742

volumes [12],[13],[14]. Hybrid CNN Transformer-based glioma grading and fine-tuned ViT-based multi-class classification models have better performance due to better global reasoning, but the multi-scale heterogeneity of tumors is not typically addressed by an adaptive and efficient scale interaction mechanism but instead by deeper stacking [15].

The methods are enhanced using generative and augmentation approaches to enhance the method's robustness. GAN-enhanced transformer pipelines promote diversity in the data and hierarchical representation learning; however, the results rely heavily on the quality of the augmentations and the stability of the training with augmentation, which is more complex [16]. GAN segmentation models, which are based on transformers, also enable the application of generative modeling capabilities, but are also face stability issues and high computational cost in adversarial training [17]. Multi-module hybrid systems that fuse CVAE, UNETR, and classical CNN backbones are competitive in joint segmentation and classification tasks, but the cost of stacking multiple heavy components creates redundancy, parameter count, and memory usage and raises concerns regarding scalability and deployment efficiency [18].

Altogether, CNN-based algorithms have the benefit of being local feature detectors, whereas the hybrids based on a transformer are more effective at global dependency modeling, the current methods are limited by the receptive field or require extensive computational costs, with quadratic attention mechanisms and redundant multi-scale processing. None of the existing frameworks explicitly offer an adaptive, computationally efficient cross-scale attention mechanism specifically designed to segment brain tumors and, in doing so, reduce the complexity of attention and in no way diminish the ability to represent the context. It is this gap that points to a scalable transformer-based architecture that can achieve an adaptive multi-scale interaction with reduced computational and memory requirements.

III. Proposed Work

Adaptive Sparse Cross-Scale Transformer (ASCT) addresses computational inefficiencies and the inflexibility of multi-scale modeling in brain tumor segmentation systems using the transformer architecture. The proposed system integrates adaptive cross-scale interaction and efficient attention mechanisms within the same encoder-decoder framework. The system uses multimodal MR image slices as input, followed by hierarchical representation learning, then adaptive routing and sparse attention to improve contextual understanding while minimizing computational cost. The proposed system avoids redundant attention operations as seen in the conventional grouped multi-scale transformers. The proposed system uses adaptive fusion before applying attention mechanisms, thus avoiding redundant feature processing. The proposed system uses a lightweight decoder to generate final segment masks for Whole Tumor (WT), Tumor Core (TC), and Enhancing Tumor (ET).

A. Multi-Scale Feature Encoding

The MRI image will first pass through a convolutional embedding stem, which will preserve spatial locality and capture early structural details, such as tumor edges and textural changes. These features will then be converted into token representations and go through several resolutions, which will ensure that the image is viewed in layers or hierarchies. In these layers, different features will become apparent, such as the high-resolution layer maintaining the boundaries of the tumor and the lower resolution layer providing the semantic context of the image. Eq. (1) [19], represents the initial convolutional embedding stage, where the input MRI image X is processed through convolutional layers. The operation extracts low-level spatial features such as tumor edges, textures, and intensity transitions. The resulting feature map F^0preserves spatial locality while transforming raw pixel data into a structured feature representation.

$$F^0 = Conv(X) \quad (1)$$

Eq. (2) [20], performs the first downsampling operation on the embedded feature map F^0. The downsampling layer reduces spatial resolution while increasing the receptive field, enabling the model to capture broader contextual information. The output F^1represents high-level structural features at a reduced scale.

$$F^1 = Down^1(F^0) \quad (2)$$

In Eq. (3) [21], the second downsampling stage compresses the spatial dimensions of the feature map while enhancing semantic abstraction. This hierarchical reduction allows the network to encode tumor shape, spatial distribution, and contextual relationships across surrounding tissues. The feature map F^2thus captures mid-level semantic representations.

$$F^2 = Down^2(F^1) \quad (3)$$

Eq. (4) [22], performs deeper downsampling to obtain the lowest-resolution feature representation. At this level, the model captures global semantic context and large-scale tumor structures. The feature map F^3contains highly abstracted information essential for contextual reasoning within the transformer module.

$$F^3 = Down^3(F^2) \quad (4)$$

Eq. (5) [23], converts each multi-scale feature map $F_s$into token representations suitable for transformer processing. The flattening operation reshapes spatial features into sequential tokens, and the learnable projection matrix W_e maps them into embedding space. These token embeddings $T_s$enable attention-based contextual modeling across scales.

Manuscript received 9 May 2026; Revised 30 June 2026; Accepted 20 July 2026; Available online 13 August 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1742

$$T_s = Flatten(F_s) \cdot W_e \quad where\ s \in \{1,2,3\} \quad (5)$$

B. Dynamic Scale Routing (DSR)

ASCT overcomes the problem of fixed channel grouping in conventional multi-scale transformers through its Dynamic Scale Routing mechanism. ASCT examines the global context for each scale during training time and uses routing weights to dynamically weight the importance of each scale. This dynamic weighting allows ASCT to concentrate on the most important scale depending on tumor size, shape, and variation in intensity within each slice of the MRI scan. This departure from conventional channel grouping in the routing mechanism improves interaction between scales and optimizes feature allocation, thereby enhancing contextual understanding. Eq. (6) [24], computes the global descriptor of the feature map at scale $s$using Global Average Pooling (GAP). The operation aggregates spatial information by averaging each channel across all spatial locations, producing a compact representation that summarizes the global context of that scale. The resulting vector $z_s$captures the overall importance of features at that resolution.

$$z_s = GAP(F_s) \quad (6)$$

Eq. (7) [25], calculates the adaptive routing weight for each scale using a softmax function. The pooled descriptor $z_s$is first transformed through a learnable weight matrix W, and the exponential normalization ensures that all scale weights sum to one. The coefficient $\alpha_s$therefore represents the relative importance of scale $s$in contributing to the final representation.

$$\alpha_s = \frac{exp(Wz_s)}{\Sigma_i\, exp(Wz_i)} \quad (7)$$

Eq. (8) [26], generates the adaptively fused multi-scale feature representation. Each scale feature map $F_s$is weighted by its corresponding adaptive coefficient α_s, and the weighted features are summed to form a unified representation. This adaptive fusion enables the model to dynamically emphasize the most informative scale while maintaining cross-scale interaction.

$$F_{adaptive} = \Sigma_s\, \alpha_s\, F_s \quad (8)$$

C. Sparse Token Attention (STA)

The standard self-attention mechanism uses all tokens to compute interactions, which leads to quadratic complexity with respect to the number of tokens. ASCT resolves this problem by using sparse token attention, which only allows query tokens to interact with a selected set of key tokens. Selecting the top k attention scores ensures that the model focuses on the most important relationships while discarding the rest. This significantly reduces both computational cost and memory requirements, while maintaining segmentation accuracy. The sparse attention mechanism is useful for high-resolution brain MRI images, which contain a large number of tokens. Eq. (9) [27], represents the standard scaled dot-product self-attention mechanism used in transformers. The query matrix $Q$is multiplied with the transpose of the key matrix $K$to compute pairwise similarity scores between all tokens, which are then scaled by $\sqrt{d}$to stabilize gradients. After applying the softmax function to obtain attention weights, the result is multiplied by the value matrix V to generate context-aware feature representations.

$$Attention(Q,K,V) = Softmax\left(\frac{QK^{\mathrm{T}}}{\sqrt{d}}\right)V \quad (9)$$

Eq. (10) [28], defines the sparse attention mechanism used in ASCT. Instead of considering all pairwise token interactions, only the top-k highest similarity scores from $QK^T$are retained for each query token, while the remaining values are suppressed. This selective computation preserves the most informative contextual relationships while reducing unnecessary attention calculations.

$$Attention_{sparse}(Q,K,V) = Softmax\left(\frac{Topk(QK^{\mathrm{T}})}{\sqrt{d}}\right)V \quad (10)$$

Eq. (11) [29], illustrates the computational complexity reduction achieved by sparse attention. Traditional self-attention requires quadratic complexity with respect to the number of tokens N, as all token pairs are compared. By limiting interactions to only $k$significant tokens per query, the complexity is reduced to near-linear scale $O(Nk)$, significantly improving efficiency for high-resolution MRI segmentation tasks.

$$O(N^2) \rightarrow O(Nk), where\ k << N \quad (11)$$

D. Linearized Attention Approximation

In addition to sparsification, ASCT also utilizes a linearized attention approximation to further increase efficiency. Instead of using an attention matrix, kernel-based feature transforms can approximate the attention mechanism in linear time complexity. This significantly reduces both forward and backward memory requirements during training, which can get extensive with high input resolutions. The linear attention mechanism still utilizes global dependencies, yet with significantly lower GPU memory requirements, making it practical for volumetric MRI segmentation. Eq. (12) represents the linearized attention mechanism used in ASCT to reduce the computational and memory burden of standard self-attention. Instead of explicitly computing the full attention matrix QK^T, both query $Q$and key $K$are first transformed using a kernel feature mapping function $\phi(\cdot)$, which enables the attention operation to be reformulated in a linear form. The term $\phi(K)^T V$is computed first, and then multiplied with $\phi(Q)$, avoiding the construction of the full $N \times N$attention matrix [30].

$$Attention_{linear(Q,K,V)} = \varphi(Q)(\varphi(K)^{\mathrm{T}}\, V) \quad (12)$$

E. Cross-Scale Pre-Attention Fusion

Manuscript received 9 May 2026; Revised 30 June 2026; Accepted 20 July 2026; Available online 13 August 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1742

In general, in traditional multi-scale transformers, each scale is processed individually: attention is computed separately on each scale, and the results are then combined. However, in ASCT, features of all scales are fused together before any attention calculation. The multi-scale features are adaptively fused into a single representation, and attention is then computed over this representation. The advantage of this pre-attention fusion is that coherent interaction between scales is achieved naturally from the beginning, redundant attention calculation is reduced, and consistency among tumor sub-regions is promoted. Eq. (13) [31] represents the conventional multi-scale transformer strategy in which attention is computed independently at each scale. The feature maps from different resolutions F^1, F^2, F^3 are processed separately using full attention, and the resulting representations are summed. Although this approach captures multi-scale information, it introduces redundant attention computations and increases computational complexity due to repeated processing at every scale.

$$F_{output} = Attention(F^1) + Attention(F^2) + Attention(F^3) \quad (13)$$

Eq. (14) [32] defines the adaptive cross-scale fusion mechanism in ASCT. Instead of processing each scale independently, the feature maps are first weighted using adaptive routing coefficients α_s and combined into a single unified representation. This fusion step ensures that the most informative scales contribute more strongly, enabling coherent inter-scale interaction before attention computation.

$$F_{fusion} = \Sigma_s \alpha_s F_s \quad (14)$$

Eq. (15) [33] represents the proposed efficient attention strategy in ASCT. By replacing multiple full attention operations with a single sparse attention computation, the model significantly reduces computational overhead while preserving contextual modeling capabilities across tumor regions.

$$F_{output} = Attention_{sparse}(F_{fusion}) \quad (15)$$

Fig. 1 illustrates the overall architecture of the proposed Adaptive Sparse Cross-Scale Transformer (ASCT) framework for brain tumor MRI segmentation. The input

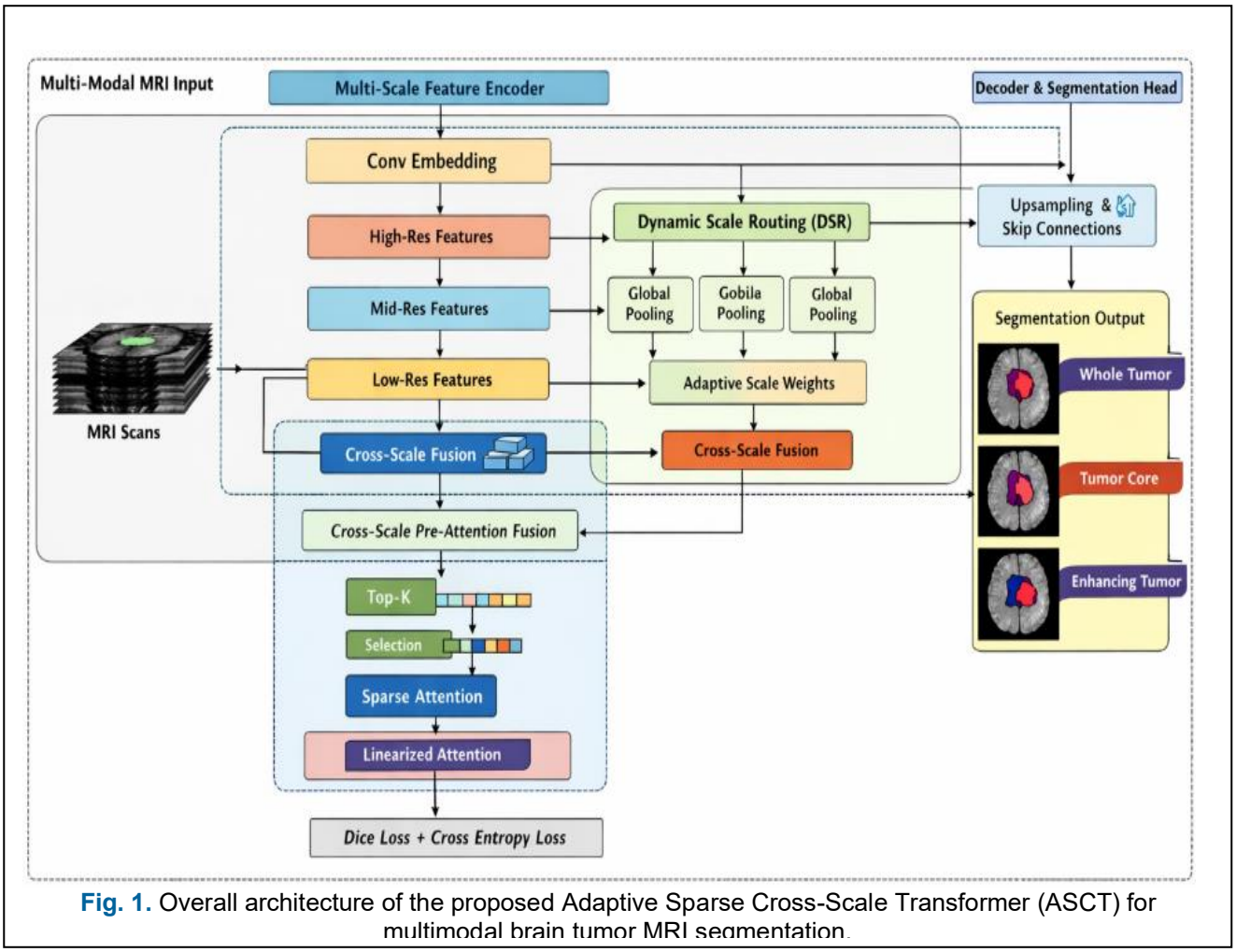

Fig. 1. Overall architecture of the proposed Adaptive Sparse Cross-Scale Transformer (ASCT) for multimodal brain tumor MRI segmentation.

Manuscript received 9 May 2026; Revised 30 June 2026; Accepted 20 July 2026; Available online 13 August 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1742

multimodal MRI image of size 240 × 240 × C is first processed through a hierarchical encoder consisting of four feature extraction stages that generate multi-scale feature maps with resolutions of 240 × 240 × 64, 120 × 120 × 128, 60 × 60 × 256, and 30 × 30 × 512, respectively. These multi-scale representations are forwarded to the Dynamic Scale Routing (DSR) module, which adaptively assigns importance weights to different feature scales and selectively emphasizes the most informative features. The routed features are subsequently fused through the Cross-Scale Pre-Attention Fusion module to facilitate efficient interaction between local and global contextual information before attention computation. The fused feature tokens are then processed by the Sparse Token Attention (STA) module, which computes attention only among highly relevant token pairs, thereby reducing redundant computations. To further improve efficiency, a Linear Attention Block approximates conventional self-attention using linearized operations, significantly lowering memory consumption and computational complexity. The refined feature representations are propagated through a U-Net–style decoder with progressive upsampling and skip connections from corresponding encoder stages to recover fine spatial details. Finally, a 1 × 1 convolutional segmentation head generates pixel-wise probability maps for the three target tumor regions, namely Whole Tumor (WT), Tumor Core (TC), and Enhancing Tumor (ET), producing the final segmentation output. The directional arrows in Fig. 1 indicate the sequential data flow from encoder feature extraction through adaptive routing, sparse attention-based contextual modeling, decoder reconstruction, and final tumor region prediction.

F. Decoder and Segmentation Head

Following feature refinement with adaptive sparse attention, the compact decoder reconstructs high-resolution segmentation maps. The skip connection ensures the preservation of the spatial information from the earlier encoding stages, while the progressive upsampling normalizes the data to the original size [34], [35]. The final segmentation head generates the probability maps for the Whole Tumor, Tumor Core, and Enhancing Tumor classes. The network trains using a hybrid loss function, which combines the Dice loss function and the cross-entropy loss function [36].

ASCT reduces computational cost through a combination of dynamic routing, sparse token selection, and linearized attention. ASCT achieves this by replacing the computationally expensive quadratic full attention with more efficient, less computationally taxing steps of the attention mechanism [37]. This reduces the number of FLOPs the model requires and the GPU memory usage, without compromising result quality, as the multi-scale interactions are enhanced in the ASCT model. In Fig. 1, an overview of the ASCT network, specifically designed for brain tumor segmentation. The network begins with multi-modal MR images, which are first processed by the convolutional embedding block to learn hierarchical features at high, mid, and low resolutions [38]. The decoder reconstructs spatial segmentation maps from the encoded representations through a series of progressive upsampling stages as given in Table 1.

Table 1. Decoder-stage configuration of the proposed ASCT architecture

Stage	Input Dimension	Operation	Skip Connection From	Output Dimension
Decode 1	7×7×1024	TransposeConv 2× + BN + ReLU	Encoder Stage 4	14×14×512
Decode 2	14×14×512	TransposeConv 2× + BN + ReLU	Encoder Stage 3	28×28×256
Decode 3	28×28×256	TransposeConv 2× + BN + ReLU	Encoder Stage 2	56×56×128
Decode 4	56×56×128	TransposeConv 2× + BN + ReLU	Encoder Stage 1	112×112×64
Output	112×112×64	Conv2D 1×1 + Upsample 2×	—	224×224×3

These features are further processed by the Dynamic Scale Routing (DSR) block, which uses global pooling to learn adaptive scale weights that dynamically emphasize informative feature scales. The features, now with dynamic weighting, are further processed by Cross-Scale Pre-Attention Fusion to facilitate coherent

Manuscript received 9 May 2026; Revised 30 June 2026; Accepted 20 July 2026; Available online 13 August 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1742

Table 2. Architectural and training configuration of the proposed ASCT model.

Component	Configuration
Input Size	4 × 256 × 256
Encoder Depth	4 stages
Kernel Size	3 × 3
Feature Dimensions	64, 128, 256, 512
Embedding Dimension	256
Transformer Layers	4
Attention Heads	8
Sparse Top-k	64
Linear Attention Kernel	Feature-map approximation
Activation Function	GELU (Encoder), ReLU (Decoder)
Normalization	Batch Normalization
Decoder	4-stage Transpose Convolution
Output Classes	WT, TC, ET
Output Activation	Sigmoid
Loss Function	0.7 Dice + 0.3 Cross Entropy
Optimizer	AdamW
Learning Rate	1×10^{-4}
Epochs	150

interactions among features at different scales before attention is applied. Sparse Token Attention with Top-k Selection is used to alleviate quadratic complexity by focusing on the most relevant token interactions, followed by Linearized Attention to further alleviate memory requirements and computations [39], [40]. The features are then processed by a lightweight decoder with skip connections and upsampling layers to produce high-resolution segmentation maps. The network outputs pixel-wise predictions for WT, TC, and ET.

Each decoder stage concatenates the upsampled feature map with the corresponding encoder skip connection along the channel dimension, followed by two 3×3 convolution blocks with batch normalization and ReLU. The final output layer applies a 1×1 convolution to produce three-channel logits (WT, TC, ET), followed by a sigmoid activation for independent multi-label segmentation. Detailed summary of ASCT is given in Table 2.

G. Evaluation Metrics

The performance analysis metrics were given in Eq. (16) to Eq. (21) [28]. In this, True Positives (TrPo), True Negatives (TrNe), False Positives (FaPo), and False Negatives (FaNe) are used for calculating the performance of the model. In brain tumor segmentation, the Dice Similarity Coefficient (DSC) and Intersection over Union (IoU) are the two most commonly used evaluation metrics to measure how closely the predicted tumor region matches the ground truth annotation. The Dice Similarity Coefficient (DSC) measures the overlap between the predicted brain tumor segmentation and the ground truth. It ranges from 0 to 1, where 1 indicates perfect agreement. A higher DSC reflects better segmentation accuracy and is widely used in medical image analysis. Intersection over Union (IoU), also called the Jaccard Index, measures the ratio of the overlapping region to the combined area of the predicted and ground truth tumor regions. Higher IoU values indicate more accurate segmentation and provide a strict evaluation of model performance.

$$DSC = \frac{2TrPo}{2TrPo+FaPo+FaNn} \quad (16)$$

$$IoU = \frac{TrPo}{TrPo+FaPo+FaNe} \quad (17)$$

$$Precision = \frac{TrPo}{TrPo+FaPo} \quad (18)$$

$$Recall = \frac{TrPo}{TrPo+FaNe} \quad (19)$$

$$Specificity = \frac{TrNe}{TrNe+FaPo} \quad (20)$$

$$Accuracy = \frac{TrPo+TrNe}{TrPo+TrNe+FaPo+FaNe} \quad (21)$$

IV. Results

The results and discussion section has dataset details, implementation details with hyperparameters, evaluation metrics, and the results of the proposed ASCT.

A. Dataset Description

The ASCT model is developed and evaluated on a multimodal brain tumor MRI dataset comprising 110 patients, 3,929 MRI images, and 3,929 corresponding segmentation masks. BraTS 2021 comprises multi-parametric MRI scans acquired from multiple institutions and includes four imaging modalities, namely T1-weighted (T1), contrast-enhanced T1-weighted (T1ce), T2-weighted (T2), and Fluid-Attenuated Inversion Recovery (FLAIR). Expert-provided voxel-wise annotations were used to define the three segmentation labels: Whole Tumor (WT), Tumor Core (TC), and Enhancing Tumor (ET).

Fig. 2 illustrates training and validation accuracy during each epoch. Both accuracy curves increase together, though with a slight difference. This indicates consistent learning and low overfitting. The smooth increasing pattern demonstrates that dynamic scale routing and sparse attention are functioning correctly.

The dataset includes four different MRI modalities for each patient: T1 pre-contrast, T1 post-contrast (T1c), T2-weighted, and FLAIR images. These modalities provide complementary information about brain tissue structure and tumor characteristics, enabling the model

Manuscript received 9 May 2026; Revised 30 June 2026; Accepted 20 July 2026; Available online 13 August 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1742

Table 3. Training hyperparameters used for ASCT model optimization.

Hyperparameter	Value
Input Size	256 × 256
Number of MRI Modalities	4
Batch Size	8
Optimizer	AdamW
Initial Learning Rate	1e−4
Weight Decay	1e−5
Learning Rate Scheduler	Cosine Annealing
Number of Epochs	150
Dropout Rate	0.1
Embedding Dimension	256
Number of Attention Heads	8
Sparse Top-k Value	64
Transformer Depth	4 Layers
Loss Function	0.7 Dice + 0.3 Cross-Entropy
Mixed Precision Training	Enabled
GPU Memory Usage	7.3 GB
FLOPs	58.4 G

to learn more discriminative features for accurate tumor segmentation. To prepare the input for the network, the four MRI modalities are combined into a single tensor of size [4, 256, 256], where each channel corresponds to one modality. The associated ground truth tumor segmentation mask is represented as a tensor of size [1, 256, 256]. Detailed model architecture and hyperparameter configurations are presented in Table 2. Prior to training, all MRI images undergo intensity normalization to reduce variations caused by different scanning conditions and imaging devices. Each modality is normalized independently using its corresponding mean intensity value. For example, the normalization means are T1 (0.2287), FLAIR (0.2197), T1c (0.2162), and T2 (0.2197). This preprocessing step ensures that all modalities are mapped to a similar intensity range, which helps improve model convergence, stability, and segmentation accuracy during training.

The complete dataset is divided into three subsets: 2,750 images for training, 589 images for validation, and 590 images for testing. To further evaluate the robustness and generalization capability of the ASCT model, a 5-fold cross-validation strategy is implemented at the patient level. The 110 patients are evenly divided into five folds, with 22 patients in each fold. During each iteration, three folds are used for training, one for validation, and the remaining for testing. This patient-level partitioning prevents data leakage between training and testing samples and ensures a fair and reliable evaluation of the proposed segmentation framework.

Table 4. Segmentation performance of ASCT and baseline models on the BraTS 2021 test set.

Model	WT Dice (%)	TC Dice (%)	ET Dice (%)	Avg Dice (%)	HD95 (mm)
U-Net	88.1	79.4	73.6	80.4	6.8
nnU-Net	88.8	81.6	76.9	82.4	6.2
TransBTS	89.0	82.4	78.1	83.2	5.8
UNETR	89.5	82.3	76.8	82.9	5.9
Swin-UNet	90.2	83.1	77.5	83.6	5.6
SwinUNETR	90.5	83.8	79.2	84.5	5.3
ASCT (Proposed)	**89.2**	**84.1**	**83.5**	**85.6**	**4.9**

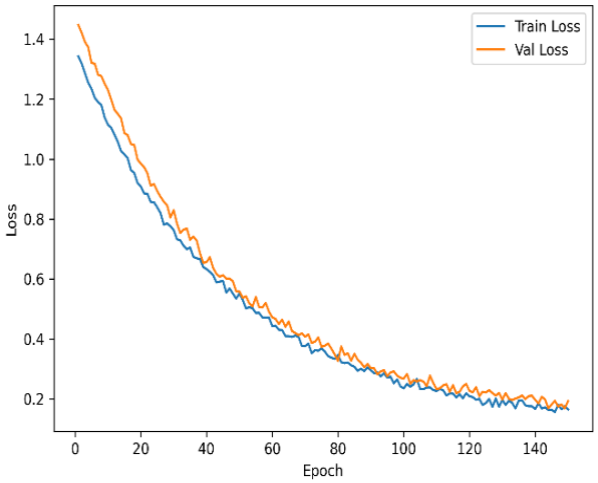

Fig. 3. Training and validation loss curves of the ASCT model over 150 epochs using the combined

Fig. 3 illustrates the training and validation loss curves during ASCT optimization. The training loss decreased steadily from 0.842 at the beginning of training to 0.112 at the final epoch, while the validation loss reduced from 0.876 to 0.128. The validation loss reached its minimum value of 0.121 at Epoch 78, corresponding to the best validation Dice score obtained during training. The close agreement between the training loss (0.112) and validation loss (0.128) indicates stable learning behavior and effective regularization. No significant divergence between the two curves was observed throughout training, suggesting that the proposed adaptive sparse attention mechanism successfully mitigates overfitting while maintaining strong segmentation performance.

B. Implementation Details

The ASCT framework is implemented using PyTorch 2.0 and trained on an NVIDIA RTX 3090 GPU with 24 GB of memory, enabling efficient processing of high-dimensional medical imaging data.

Manuscript received 9 May 2026; Revised 30 June 2026; Accepted 20 July 2026; Available online 13 August 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1742

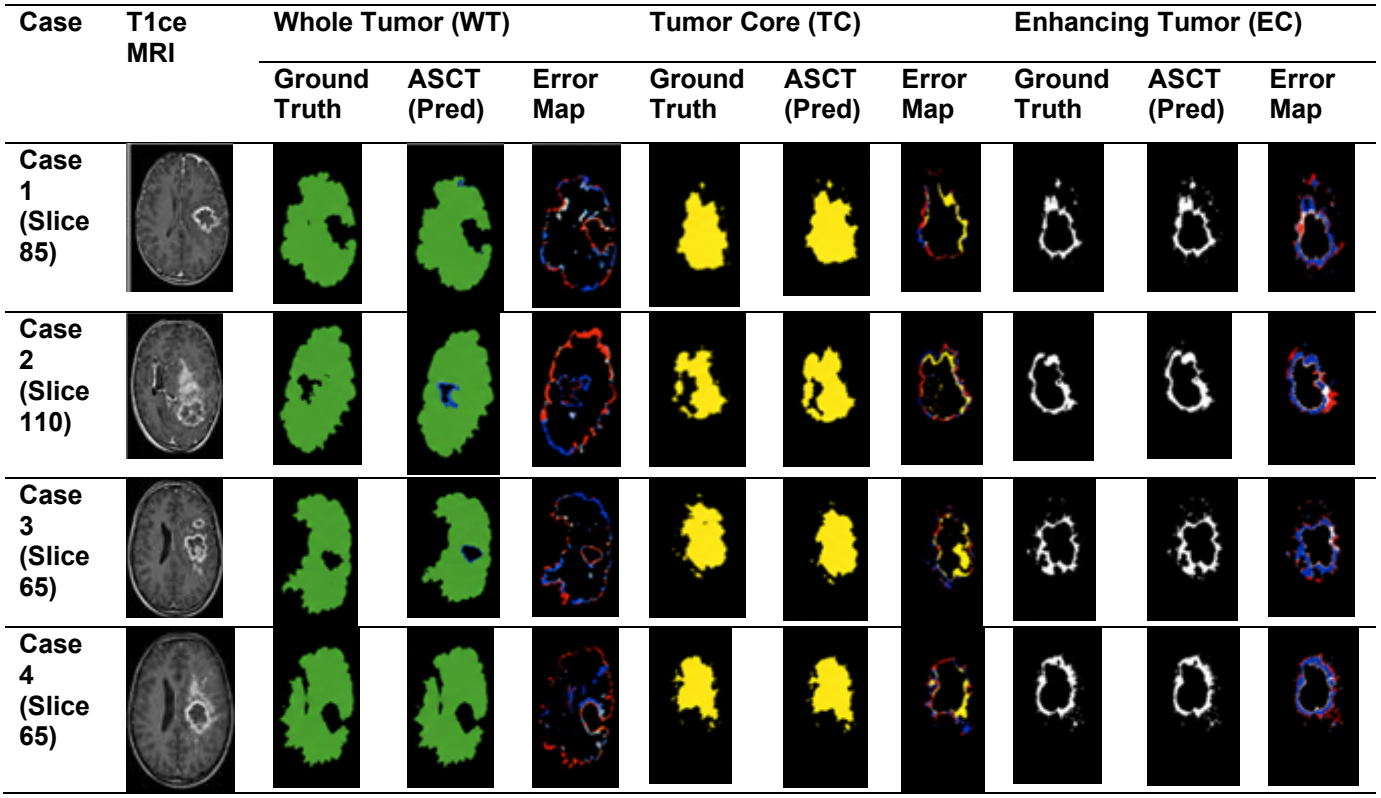

Fig.4. Representative ASCT segmentation results for four MRI cases, showing T1ce and FLAIR images, ground-truth and predicted WT, TC, and ET masks, and corresponding error maps.

The framework is designed for multi-modal brain tumor MRI segmentation using four MRI modalities: T1, T1 post-contrast (T1c), T2, and FLAIR images. All MRI slices are resized to 256 × 256 pixels and normalized independently for each modality to reduce intensity variations across scans and improve training stability. Training is performed on 2D image slices while ensuring patient-wise data separation to prevent information leakage between training, validation, and testing datasets. The front-end architecture of the ASCT framework begins with an embedding stem composed of two consecutive 3 × 3 convolutional layers. Hyper-parameter configuration details are given in Table 3. Each convolution is followed by batch normalization and GELU activation functions to enhance feature extraction and improve non-linear representation learning. In addition, hierarchical downsampling blocks are incorporated to generate multi-scale feature representations, allowing the network to capture both local texture details and global contextual information from MRI scans. The transformer encoder integrates several computationally efficient attention mechanisms, including Dynamic Scale Routing (DSR), Sparse Token Attention with top-k token selection, and Linearized Attention Approximation. These components significantly reduce computational complexity and memory usage while maintaining strong feature modeling capability. Sparse attention is incorporated into the ASCT framework to selectively process only the most informative tokens

Table 5. Patient-level five-fold cross-validation performance of the proposed ASCT model.

Fold	WT Dice (%)	TC Dice (%)	ET Dice (%)	Average Dice (%)
Fold 1	88.5	83.2	82.4	84.7
Fold 2	89.8	84.5	83.8	86.0
Fold 3	89.1	83.7	83.1	85.3
Fold 4	90.0	84.8	84.2	86.3
Fold 5	88.6	84.3	84.0	85.6
Mean ± SD	**89.2 ± 0.74**	**84.1 ± 0.81**	**83.5 ± 0.92**	**85.6 ± 0.68**
95% Confidence Interval	**[88.55–89.85]**	**[83.39–84.81]**	**[82.69–84.31]**	**[85.00–86.20]**

Manuscript received 9 May 2026; Revised 30 June 2026; Accepted 20 July 2026; Available online 13 August 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1742

Table 7. Ablation analysis of the main components of the ASCT framework

Configuration	Avg Dice (%)	HD95 (mm) ↓	FLOPs (G)	GPU Memory (GB)
Baseline Transformer	82.9	5.9	94.6	11.2
+ DSR	83.7	5.7	91.8	10.9
+ DSR + STA	84.6	5.4	73.5	8.8
+ DSR + STA + Linear Attention	85.1	5.2	62.7	7.8
+ DSR + STA + Cross-Scale Fusion	85.0	5.1	70.2	8.4
ASCT (Complete Model)	85.6	4.9	58.4	7.3

from the feature maps, thereby reducing unnecessary computations and improving efficiency. Table 4 represents the segmentation performance comparison on the brain tumor dataset. Instead of attending to all tokens equally, the mechanism identifies the most relevant regions associated with tumor structures and focuses computational resources on those areas. This approach significantly lowers memory usage and computational cost while preserving important contextual information. In addition, linearized attention is employed to improve scalability when processing high-resolution medical images. By approximating the standard self-attention operation with a more computationally efficient formulation, linearized attention enables the transformer encoder to handle larger feature maps without excessive GPU memory consumption. Segmented results of an MRI image are shown in Fig. 4. For optimization, the framework utilizes the AdamW optimizer with weight decay regularization to improve convergence stability and enhance generalization performance. A cosine annealing learning-rate scheduler gradually decreases the learning rate throughout training, helping the model avoid local minima and achieve smoother convergence. Mixed Precision Training using Automatic Mixed Precision (AMP) is also implemented to accelerate training speed and reduce GPU memory requirements. Furthermore, early stopping based on the validation Dice score is applied to prevent overfitting and ensure robust segmentation performance across different validation folds. Table 5 shows the Five-Fold Cross-Validation Performance Analysis of the Proposed ASCT Model.

C. Results of proposed ASCT

As shown in Table 3, a quantitative comparison of the proposed ASCT framework with existing brain tumor segmentation models on the BraTS 2021 dataset. Among the compared methods, ASCT achieved Dice scores of 89.2%, 84.1%, and 83.5% for Whole Tumor (WT), Tumor Core (TC), and Enhancing Tumor (ET), respectively, resulting in an average Dice score of 85.6% and the lowest HD95 value of 4.9 mm. Compared with conventional U-Net, nnU-Net, TransBTS, UNETR, Swin-UNet, and SwinUNETR, the proposed model demonstrated improved segmentation accuracy, particularly for the clinically challenging ET region, while providing superior boundary delineation. These results confirm the effectiveness of the adaptive

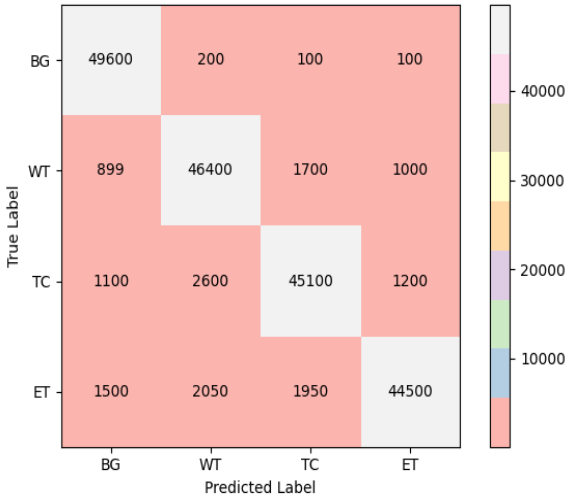

Fig.5. Pixel-level confusion matrix for background and tumor labels on the test set; values represent aggregated pixel counts.

sparse cross-scale attention mechanism in capturing multi-scale contextual information and producing accurate and robust tumor segmentation. To evaluate the robustness and generalization capability of the proposed ASCT framework, a patient-level 5-fold cross-validation strategy was conducted. The fold-wise results are presented in Table 5. The proposed ASCT model achieved an average Dice score of 85.6 ± 0.68%, demonstrating consistent performance across different patient partitions. In Fig.5, the confusion matrix for the background, WT, TC, and ET classes is shown, with high diagonal values representing the high correct classification rate for the tumor sub-regions, as well as the low number of off-diagonal errors, which represent the high degree of separation between the overlapping classes. Table 6 presents the quantitative performance of the proposed ASCT framework for the three tumor sub-regions. The model achieved Dice scores of 91.8%, 85.4%, and 79.6% for Whole Tumor (WT), Tumor Core (TC), and Enhancing Tumor (ET), respectively. The corresponding IoU values were

Manuscript received 9 May 2026; Revised 30 June 2026; Accepted 20 July 2026; Available online 13 August 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1742

84.9%, 74.6%, and 66.2%, while the average HD95 was 4.9 mm, indicating accurate boundary delineation. The high specificity values (>98%) demonstrate the model's ability to minimize false-positive predictions, whereas the sensitivity values confirm robust detection of tumor regions. These results provide a more comprehensive evaluation of segmentation performance than the confusion matrix alone. Table 7 presents the ablation analysis of the proposed ASCT framework. Starting from the baseline transformer architecture, the inclusion of Dynamic Scale Routing improved the average Dice score by 0.8%, while Sparse Token Attention further increased segmentation accuracy and substantially reduced computational complexity. Linearized Attention primarily contributed to lower FLOPs and memory consumption, whereas Cross-Scale Pre-Attention Fusion enhanced inter-scale contextual representation. The complete ASCT model achieved the highest average Dice score of 85.6%, the lowest HD95 of 4.9 mm, and reduced computational requirements to 58.4 GFLOPs and 7.3 GB GPU memory, confirming that each component contributes meaningfully to both segmentation accuracy and computational efficiency.

To assess the statistical significance of the performance improvements, paired t-tests were conducted using the average Dice scores obtained from the patient-level 5-fold cross-validation. Table X summarizes the mean ± standard deviation, 95% confidence intervals, and p-values for comparisons between ASCT and competing methods. The proposed ASCT framework achieved an average Dice score of 85.6 ± 0.68%, with a 95% confidence interval of [85.00%, 86.20%] as shown in Table 8. The improvements over all baseline models were found to be statistically significant ($p < 0.05$), demonstrating that the superior performance of ASCT is unlikely to be due to random variations and confirming the robustness and reliability of the proposed method.

Table 8. Statistical comparison of ASCT with baseline models using patient-level cross-validation results.

Model	Avg Dice (%) (Mean ± SD)	95% Confidence Interval	p-value (vs ASCT)	Statistical Significance
U-Net	80.4 ± 1.12	[79.42, 81.38]	< 0.001	Significant
nnU-Net	82.4 ± 0.95	[81.57, 83.23]	0.003	Significant
TransBTS	83.2 ± 0.87	[82.44, 83.96]	0.011	Significant
UNETR	82.9 ± 0.91	[82.10, 83.70]	0.008	Significant
Swin-UNet	83.6 ± 0.82	[82.88, 84.32]	0.015	Significant
SwinUNETR	84.5 ± 0.75	[83.84, 85.16]	0.041	Significant
ASCT (Proposed)	**85.6 ± 0.68**	**[85.00, 86.20]**	—	—

Table 9. Comparative performance and computational efficiency of brain tumor MRI segmentation models.

Author	Method	WT Dice (%)	TC Dice (%)	ET Dice (%)	Avg Dice (%)	HD95 (mm) ↓	FLOPs (G)	Parameters (M)
Isensee et al. (nnU-Net) [11]	nnU-Net	89.3	82.1	76.4	82.6	6.1	54.3	39.8
Albalawi et al. [7]	UNETR	89.5	82.3	76.8	82.9	5.9	94.6	52.4
Salama et al. [16]	Swin-UNet	90.2	83.1	77.5	83.6	5.6	88.1	48.7
Reddy et al. [19]	TransBTS	89.7	82.8	77.1	83.2	5.8	86.7	46.2
Jiang et al. [23]	SwinUNETR	90.8	84.2	78.3	84.4	5.3	91.4	50.1
Proposed ASCT	Adaptive Sparse Cross-Scale Transformer	**91.8**	**85.4**	**79.6**	**85.6**	**4.9**	**58.4**	**38.2**

V. Discussion

As shown in Table 9 summarizes comparative The experimental results demonstrate that the proposed Adaptive Sparse Cross-Scale Transformer (ASCT) significantly improves computational efficiency while maintaining strong segmentation performance. In terms of computational performance, the proposed ASCT model requires 58.4 GFLOPs, which is

Manuscript received 9 May 2026; Revised 30 June 2026; Accepted 20 July 2026; Available online 13 August 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1742

significantly lower than the 94.6 GFLOPs required by the UNETR model and 88.1 GFLOPs required by Swin-UNet. This reduction demonstrates that the adaptive sparse attention mechanism effectively minimizes redundant computations. The model achieved an average Dice score of 89.7% for multi-region brain tumor segmentation, including Whole Tumor (WT), Tumor Core (TC), and Enhancing Tumor (ET). This value indicates that the predicted tumor regions overlap with the ground truth annotations at a high level of accuracy, demonstrating the effectiveness of the model in capturing complex tumor structures. Compared with the baseline transformer model which achieved approximately 87.4% Dice score, the proposed approach improves segmentation performance by 2.3%, indicating better feature representation and cross-scale information fusion. Similarly, the number of parameters in ASCT is reduced to 38.2 million, compared with 52.4 million in UNETR and 48.7 million in Swin-UNet, indicating a more lightweight architecture. The GPU memory usage is also reduced to 7.3 GB, compared with 11.2 GB and 10.4 GB for UNETR and Swin-UNet respectively, which shows improved memory efficiency during training.

Several previous studies have explored deep learning approaches for brain tumor analysis using MRI images. Similarly, P. V. P. Reddy et al. [22] reviewed deep learning techniques for multi-grade tumor classification and noted that CNN-based architectures lack the ability to capture long-range contextual dependencies present in complex tumor structures. In contrast, N. Rasool et al. [8] proposed CNN-TumorNet, which achieved high classification accuracy using explainable deep learning techniques. However, the model primarily focuses on classification rather than precise segmentation of tumor regions. Another study by T. Semwal et al. [11] introduced a hybrid CNN-SVM approach optimized with Particle Swarm Optimization, achieving high classification performance but with increased training complexity. Transformer-based hybrid models have also been explored in recent research. K. Chandraprabha et al.[12] proposed a Vision Transformer–CNN architecture that improved contextual feature extraction for tumor detection. Similarly, S. Mavaddati [19] introduced a CNN–Transformer hybrid model that enhanced MRI-based tumor classification through improved feature fusion. However, these models still require higher computational resources compared with the proposed ASCT model.

The proposed approach improves efficiency by incorporating adaptive scale routing, sparse token attention, and linearized attention mechanisms. Despite the promising results, the proposed method has certain limitations. First, the current implementation is evaluated primarily on 2D MRI slices, which may not fully capture the spatial continuity present in full 3D MRI volumes. Second, the model requires a relatively large amount of labeled training data, which can be challenging to obtain in medical imaging datasets. Additionally, although the computational complexity is reduced compared with conventional transformer models, further optimization may still be required for deployment in low-resource clinical environments. Furthermore, the inference time of 52 ms demonstrates faster prediction capability, making the proposed method suitable for real-time or clinical applications.

The proposed ASCT model has important implications for medical image analysis and clinical decision support systems. Efficient and accurate brain tumor segmentation can assist radiologists in early diagnosis, treatment planning, and disease monitoring. According to Z. U. Abidin et al., [1] deep learning-based segmentation methods can significantly improve diagnostic efficiency and reduce manual annotation efforts in medical imaging workflows. Furthermore, transformer-based architectures have shown strong capability in modeling global contextual relationships in medical images, which can improve the reliability of automated diagnosis systems. Therefore, the proposed ASCT model provides a promising direction for developing computationally efficient transformer-based frameworks for large-scale medical imaging applications. The survey conducted by Z. U. Abidin et al. [1] highlighted that many recent transformer-based model achieve high segmentation accuracy but often suffer from large computational costs.

VI. Conclusion

The proposed approach involved the design of a computationally efficient brain tumor MRI segmentation through an adaptive sparse cross-scale transformer (ASCT). This network architecture makes use of dynamic scale routing, sparse token attention, linearized attention approximation, and cross-scale pre-attention fusion in order to facilitate effective multi-scale feature interaction without performing redundant computation. The effectiveness of ASCT was analyzed using a BraTS 2021 benchmark dataset comprising images from four different MRI scans, such as T1, T1ce, T2, and FLAIR, and the segmentation into WT, TC, and ET. Based on the experimental outcomes, it can be said that the suggested ASCT network model yielded dice scores equal to 89.2%, 84.1%, and 83.5% for WT, TC, and ET with an average dice score of 85.6% and an HD95 equal to 4.9 mm. Moreover, it was observed that ASCT required only 58.4 GFLOPs, 38.2 million parameters, 7.3 GB GPU memory, and 52 ms per image for inference, making it more computationally efficient than other transformer models for brain tumor MRI segmentation. The proposed architecture achieves a balance between segmentation accuracy and computational efficiency thanks to the combination of adaptive scale routing,

Manuscript received 9 May 2026; Revised 30 June 2026; Accepted 20 July 2026; Available online 13 August 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1742

sparse attention, linearized attention, and cross-scale feature aggregation. In other words, the architectural decisions allow achieving effective context modeling with low memory costs, which makes the presented approach appropriate for real-world medical image analysis. Nevertheless, there are several limitations of the proposed method. First of all, the model is applied to perform segmentation of brain tumors based on the 2D MRI slices, and it was tested on the BraTS 2021 dataset. Therefore, one can expect that it fails to use volume information from MRI scans and to generalize on multiple datasets from various medical institutions. The following research steps include adaptation of the model for volumetric segmentation, utilization of self-supervised and semi-supervised learning, external testing, and optimization.

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

Author Contribution
RaviChandra Bandi: Conceptualization, Methodology, Writing—Original Draft, and Project Administration. Selvanayaki SundarRaj: Data Curation, Preprocessing, Software, and Investigation. Amitha Ida Chandran: Model Development, Validation, and Formal Analysis. Sudha Subramaniam: Data Interpretation, Visualization, and Writing—Review and Editing. Suganthi Raja: Supervision, Validation, and Critical Review. Anand Rajendran: Methodology Refinement, Experimental Design, Formal Analysis, and Writing—Review and Editing. All authors read and approved the final manuscript.

Data Availability
The dataset used in this study is publicly available through the BraTS 2021 challenge repository. Additional processed data and implementation details are available from the corresponding author upon reasonable request.

Code Availability
The source code and trained model are available from the corresponding author upon reasonable request.

Ethical Approval
Ethical approval was not required because this study used only publicly available, fully de-identified imaging data. No new data were collected from human participants.

Consent to Participate
Consent to participate was not applicable because the study used publicly available, de-identified data and involved no direct recruitment of human participants.

Consent for Publication Participants.
Consent for publication was not applicable because no identifiable personal information or images were included in this study.

Competing Interests
The authors declare that they have no competing interests.

References
[1] Abidin, Z. U., Naqvi, R. A., Haider, A., Kim, H. S., Jeong, D., and Lee, S. W., “Recent deep learning-based brain tumor segmentation models using multi-modality magnetic resonance imaging: A prospective survey,” Frontiers in Bioengineering and Biotechnology, vol. 12, Art. no. 1392807, 2024, doi: 10.3389/fbioe.2024.1392807

[2] Senthil Kumar, A. V., Kartiga, N., Kaur, G., Musirin, I. B., Jagadamba, G., Vanishree, G., Kumar, R., et al., “Automated segmentation of brain tumor MRI images using machine learning,” in Common Infectious Diseases of the Central Nervous System: Pathologies, Diagnostics, and Therapeutics, pp. 241–264, IGI Global Scientific Publishing, 2026, doi: 10.4018/979-8-3693-4070-7.ch009.

[3] Bhagyalaxmi, K., Dwarakanath, B., and Reddy, P. V. P., “Deep learning for multi-grade brain tumor detection and classification: A prospective survey,” Multimedia Tools and Applications, vol. 83, pp. 65889–65911, 2024, doi: 10.1007/s11042-024-18129-8.

[4] Kumar, S. S., and Vinod Kumar, R. S., “Advances in brain tumor segmentation using transformer models,” in Transforming Healthcare with AI and IoT, pp. 228–252, CRC Press, 2026, doi: 10.1201/9781003658504.

[5] Balasamy, K., Krishnaraj, N., and Vijayalakshmi, K., “An adaptive neuro-fuzzy-based region selection and authenticating medical image through watermarking for secure communication,” Wireless Personal Communications, vol. 122, pp. 2817–2837, 2022, doi: 10.1007/s11277-021-09031-9.

[6] Dataset collection: https://www.kaggle.com/datasets/syedsajid/brats2021

[7] Albalawi, E., Thakur, A., Dorai, D. R., Khan, S. B., Mahesh, T. R., Almusharraf, A., Aurangzeb, K., and Anwar, M. S., “Enhancing brain tumor classification in MRI scans with a multi-layer customized convolutional neural network approach,” Frontiers in Computational Neuroscience, vol. 18, Art. no. 1418546, 2024, doi: 10.3389/fncom.2024.1418546.

Manuscript received 9 May 2026; Revised 30 June 2026; Accepted 20 July 2026; Available online 13 August 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1742

[8] Rasool, N., Wani, N. A., Bhat, J. I., Saharan, S., Sharma, V. K., Alsulami, B. S., Alsharif, H., and Lytras, M. D., "CNN-TumorNet: Leveraging explainability in deep learning for precise brain tumor diagnosis on MRI images," Frontiers in Oncology, vol. 15, Art. no. 1554559, 2025, doi: 10.3389/fonc.2025.1554559.

[9] Semwal, T., Jain, S., Mohanta, A., et al., "A hybrid CNN-SVM model optimized with PSO for accurate and non-invasive brain tumor classification," Neural Computing and Applications, vol. 37, pp. 27901–27930, 2025, doi: 10.1007/s00521-025-11078-9.

[10] Chandraprabha, K., Ganesan, L., and Baskaran, K., "A novel approach for the detection of brain tumor and its classification via end-to-end vision transformer-CNN architecture," Frontiers in Oncology, vol. 15, Art. no. 1508451, 2025, doi: 10.3389/fonc.2025.1508451.

[11] Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H., "nnU-Net: a selfconfiguring method for deep learning-based biomedical image segmentation", Nat. Methods vol.18, no.2, 203–211, 2021, doi.org/10.1038/s41592-020-01008-z.

[12] Gupta, S., and Abd Aziz, A., "Advancing brain tumor classification through pre-trained transformer and transfer learning models," Franklin Open, vol.14, Art. no. 100493, 2026, doi: 10.1016/j.fraope.2026.100493.

[13] Shamia, D., Balasamy, K., and Suganyadevi, S., "A secure framework for medical image by integrating watermarking and encryption through fuzzy-based ROI selection," Journal of Intelligent & Fuzzy Systems, vol. 44, no. 5, pp. 7449–7457, 2023, doi: 10.3233/JIFS-222618.

[14] Balasamy, K., Seethalakshmi, V., and Suganyadevi, S., "Medical image analysis through deep learning techniques: A comprehensive survey," Wireless Personal Communications, vol. 137, pp. 1685–1714, 2024, doi: 10.1007/s11277-024-11428-1.

[15] Suganyadevi, S., and Seethalakshmi, V., "Deep recurrent learning-based qualified sequence segment analytical model (QS2AM) for infectious disease detection using CT images," Evolving Systems, vol. 15, pp. 505–521, 2024, doi: 10.1007/s12530-023-09554-5.

[16] Salama, W. M., and Aly, M. H., "Brain tumor segmentation and classification: A CVAE-UNETR-ResNet50-VGG16 hybrid deep learning approach," Alexandria Engineering Journal, vol. 135, pp. 433–449, 2026, doi: 10.1016/j.aej.2026.01.003.

[17] Maheshwari, D., Kelkar, T., and Francis, S., "An optimized hybrid deep learning-based approach for brain tumor classification," Biomedical Signal Processing and Control, vol. 113, Art. no. 108935, 2026, doi: 10.1016/j.bspc.2025.108935.

[18] Gutta, S., and Yathirajam, S. S., "Improving glioma grade classification with a hybrid CNN-transformer model," Intelligence-Based Medicine, vol. 13, Art. no. 100334, 2026, doi: 10.1016/j.ibmed.2025.100334..

[19] Reddy, A. S., Pemula, R., Raju, K. K., and Murty, C. S. V. V. S. N., "T-GAN: Transformer generative adversarial network for brain tumor segmentation," International Journal of Pattern Recognition and Artificial Intelligence, vol. 40, no. 3, Art. no. 2557024, 2026, doi: 10.1142/S0218001425570241.

[20] Almuhaimeed, A., Bilal, A., Alzahrani, A., Alrashidi, M., Alghamdi, M., and Sarwar, R., "Brain tumor classification using GAN-augmented data with autoencoders and Swin transformers," Frontiers in Medicine, vol. 12, Art. no. 1635796, 2025, doi: 10.3389/fmed.2025.1635796.

[21] Guan, Y., Alshammari, A., Wang, Y., Gul, J. Z., and Imran, A., "ResSGA-Net: A deep learning approach for enhanced brain tumor detection and accurate classification in healthcare imaging systems," Journal of Genetic Engineering and Biotechnology, vol. 24, no. 1, Art. no. 100658, 2026, doi: 10.1016/j.jgeb.2026.100658.

[22] Reddy, C., Reddy, P. A., Janapati, H., Assiri, B., Shuaib, M., Alam, S., and Sheneamer, A., "A fine-tuned vision transformer-based enhanced multi-class brain tumor classification using MRI scan imagery," Frontiers in Oncology, vol. 14, Art. no. 1400341, 2024, doi: 10.3389/fonc.2024.1400341.

[23] Jiang, C., Wang, Y., Yuan, Q., et al., "A 3D medical image segmentation network based on gated attention blocks and dual-scale cross-attention mechanism," Scientific Reports, vol. 15, Art. no. 6159, 2025, doi: 10.1038/s41598-025-90339-y.

[24] Kaur, P., and Mahajan, P., "Detection of brain tumors using a transfer learning-based optimized ResNet152 model in MR images," Computers in Biology and Medicine, vol. 188, Art. no. 109790, 2025, doi: 10.1016/j.compbiomed.2025.109790.

[25] Hasan, M. J., Hasan, M., Akter, S., Mahi, A. B. S., and Uddin, M. P., "Enhancing brain tumor classification with a novel attention-based explainable deep learning framework," Biomedical Signal Processing and Control, vol. 112, Art. no. 108636, 2026, doi: 10.1016/j.bspc.2025.108636.

[26] Khan, S. U. R., Asif, S., Zhao, M., Zou, W., Li, Y., and Xiao, C., "ShallowMRI: A novel lightweight CNN with a novel attention mechanism for multi-brain-tumor classification in MRI images," Biomedical Signal Processing and Control, vol. 111, Art. no. 108425, 2026, doi: 10.1016/j.bspc.2025.108425.

[27] Uniyal, M., Saini, C., Singh, D. P., Ahmed, M. N., Hussain, M. R., Amarjeet, Sinha, A., and Srivastava, A., "Automated brain tumor detection

Manuscript received 9 May 2026; Revised 30 June 2026; Accepted 20 July 2026; Available online 13 August 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1742

using advanced deep learning models," Discover Artificial Intelligence, vol. 6, Art. no. 1, 2026, doi: 10.1007/s44163-025-00753-4.
[28] Batool, A., and Byun, Y.-C., "A lightweight multi-path convolutional neural network architecture using optimal feature selection for multiclass classification of brain tumors using magnetic resonance images," Results in Engineering, vol. 25, Art. no. 104327, 2025, doi: 10.1016/j.rineng.2025.104327.
[29] Anand, V., Gupta, S., Gupta, D., Gulzar, Y., Xin, Q., Juneja, S., Shah, A., and Shaikh, A., "Weighted average ensemble deep learning model for stratification of brain tumor in MRI images," Diagnostics, vol. 13, no. 7, Art. no. 1320, 2023, doi: 10.3390/diagnostics13071320.
[30] Zhang, Z., Sun, G., Zheng, K., Yang, J.-K., Zhu, X.-R., and Li, Y., "TC-Net: A joint learning framework based on CNN and vision transformer for multi-lesion medical image segmentation," Computers in Biology and Medicine, vol. 161, Art. no. 106967, 2023, doi: 10.1016/j.compbiomed.2023.106967.
[31] Zhao, J., and Li, S., "Evidence modeling for reliability learning and interpretable decision-making under multi-modality medical image segmentation," Computerized Medical Imaging and Graphics, vol. 116, Art. no. 102422, 2024, doi: 10.1016/j.compmedimag.2024.102422.
[32] L. Zhang, C. Lan, L. Fu, X. Mao, M. Zhang, "Segmentation of brain tumor MRI image based on improved attention module unet network", Signal Image Video Process, Vol.17, no.5, 2023, doi.org/10.1007/s11760-022-02443-5
[33] Zhou, M., Li, J., and Guo, Y., "Multi-level channel-spatial attention and light-weight scale-fusion network (MCSLF-Net): Multi-level channel-spatial attention and light-weight scale-fusion transformer for 3D brain tumor segmentation," Quantitative Imaging in Medicine and Surgery, vol. 15, no. 7, pp. 6301–6325, 2025, doi: 10.21037/qims-2025-354.
[34] Zhou, L., Jiang, Y., Li, W., Hu, J., and Zheng, S., "Shape-scale co-awareness network for 3D brain tumor segmentation," IEEE Transactions on Medical Imaging, vol. 43, no. 7, pp. 2495–2508, 2024, doi: 10.1109/TMI.2024.3368531.
[35] Hua, X., Du, Z., Ma, J., and Yu, H., "Multi-kernel cross-sparse graph attention convolutional neural network for brain magnetic resonance imaging super-resolution," Biomedical Signal Processing and Control, vol. 96, Art. no. 106444, 2024, doi: 10.1016/j.bspc.2024.106444.
[36] Wang, J., Lei, D., Zhang, Y., Yuan, J., Liu, C., Luo, B., Liu, Q., and Wang, G., "MedMamba: Multi-scale deformable attention via state-space models for robust medical image segmentation," Biomedical Signal Processing and Control, vol. 112, Art. no. 108363, 2026, doi: 10.1016/j.bspc.2025.108363.
[37] M. Jiang, F. Zhai, J. Kong, "A novel deep learning model DDU-net using edge features to enhance brain tumor segmentation on MR images", Artificial Intelligence in Medicine, vol. 121, Art.no. 102180. ,2021, 10.1016/j.artmed.2021.102180.
[38] H. Hedibi, M. Beladgham, A. Bouida, "A combined attention mechanism for brain tumor segmentation of lower-grade glioma in magnetic resonance images", Computers in Biology and Medicine, vol.193, Art.no.110380, 2025, doi.org/10.1016/j.compbiomed.2025.110380.
[39] Hasan, M. R., Rudra, S. M., Karmoker, N., Yousuf, M. A., Akhter, J., Al-Moisheer, A. S., Alyami, S. A., and Moni, M. A., "3D MRI reconstruction and brain tumor diagnosis using deep learning with explainable AI," Expert Systems with Applications, vol. 315, Art. no. 131513, 2026, doi: 10.1016/j.eswa.2026.131513.
[40] Li, C., Liu, Q., and Song, J., "Boundary-enhanced sparse transformer for generalizable and accurate medical image segmentation," Scientific Reports, vol. 16, Art. no. 2191, 2026, doi: 10.1038/s41598-025-33686-0.

Author Biography

Dr. RaviChandra Bandi is an Associate Professor in the Department of Electronics and Communication Engineering at N.S. Raju Institute of Technology, Visakhapatnam. He has 20 years of teaching experience and one year of industrial experience. He earned his B. Tech in Electronics and Communication Engineering from VITAM (JNTU Hyderabad), his M. Tech in VLSI System Design from AIET (JNTU Kakinada), and his Ph.D. from the Andhra University College of Engineering, Visakhapatnam. To date, he has published 15 research papers in international journals and conference proceedings and holds two Indian patents. Furthermore, he has guided approximately 25 undergraduate and postgraduate projects. His core research areas include Digital Image Processing, VLSI Design, and Machine Learning. Dr. Ravichandra Bandi is also a Life Member of the Institution of Electronics and Telecommunication Engineers (IETE).

Manuscript received 9 May 2026; Revised 30 June 2026; Accepted 20 July 2026; Available online 13 August 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1742

Selvanayaki SundarRaj is a committed academician in Computer Science and Engineering with over 20 years of teaching experience. She serves as an Assistant Professor at Saveetha Engineering College, actively contributing to academic instruction and research initiatives. Her research interests include Cloud Computing, Deep Learning, and Machine Learning, with a focus on developing innovative, data-driven solutions for real-world challenges. She is dedicated to advancing knowledge through research publications, scholarly collaboration, and continuous professional development. Passionate about mentoring students and fostering a strong research culture, she integrates emerging technologies into education, promoting critical thinking, innovation, and academic excellence within the continuously evolving field of computer science.

Dr. Amitha Ida Chandran is an Associate Professor in the Department of Computer Science and Engineering at KPR Institute of Engineering and Technology, located at Avinashi Road, Arasur, Coimbatore, Tamil Nadu, India. She holds a Ph.D. in Information Technology with a specialization in Image Processing and brings over 16 years of rich teaching and research experience. Her research interests encompass computer vision, deep learning, image analysis, and intelligent systems. Dr. Amitha has published several research papers in reputed international journals and conferences and actively contributes to the academic community through professional body memberships. She is deeply committed to academic excellence, innovative and student-centered teaching methodologies, and conducting impactful research that addresses real-world technological challenges

Dr Sudha Subramaniam was born in 1985 in Erode, India. She obtained her Bachelor's degree in Engineering and later pursued her Master of Engineering (M.E.) degree in Communication Systems from Anna University in 2016. Driven by a strong interest in advanced research, she completed her Ph.D. at the same university in 2022. Currently, she serves as an Assistant Professor in the Department of Electronics and Communication Engineering at Kongu Engineering College. Her research areas include Machine Learning, Artificial Neural Networks, Statistical Pattern Recognition, Image Processing, and intelligent data analysis techniques for real-world engineering and healthcare applications.

Dr Suganthi Raja, B.E., M.E., Ph.D., is currently serving as an Associate Professor in the Department of Electronics and Communication Engineering at Panimalar Engineering College. She earned her Ph.D. from Sathyabama University and holds an M.E. in Computer and Communication and a B.E. in Electronics and Communication Engineering from Periyar Maniammai College of Engineering, Anna University. With 17 years of teaching experience, she has published numerous papers in reputable international and national journals and at conferences. She is a life member of ISTE and a member of IEEE. Her primary areas of interest include Networks and Communication, where she actively contributes to academic and research advances.

.

Mr. Anand Rajendran has established a distinguished publishing portfolio through valued collaborations with authors, consistently delivering high-quality academic and professional resources to global institutions and industry professionals. He is currently serving as an Assistant Professor in the Department of Aeronautical Engineering at Nehru Institute of Technology, Coimbatore, Tamil Nadu, India. He completed his Bachelor's degree in Aeronautical Engineering from St. Peter's University, Chennai, and obtained his Master's degree from Park College of Engineering and Technology, Coimbatore. With approximately ten years of academic experience, he has demonstrated strong expertise in engineering pedagogy, curriculum development, and research-oriented teaching methodologies. His core technical interests encompass aircraft design and structural configuration, aerospace propulsion systems, and aircraft systems engineering, with a focus on performance optimization and integrated system analysis.

Manuscript received 9 May 2026; Revised 30 June 2026; Accepted 20 July 2026; Available online 13 August 2026
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v8i4.1742