IEEE Open Journal of the Computer Society

Received 23 January 2025; revised 7 May 2025; accepted 11 May 2025. Date of publication 12 May 2025; date of current version 3 June 2025. The review of this article was arranged by Associate Editor Juan Domingo Velasquez.

Digital Object Identifier 10.1109/OJCS.2025.3569758

TuSegNet: A Transformer-Based and Attention-Enhanced Architecture for Brain Tumor Segmentation

MIR NAFIUL NAGIB[1], RAHAT PERVEZ[2], AFSANA ALAM NOVA[1], HADIUR RAHMAN NABIL[3], ZEYAR AUNG[4] (Senior Member, IEEE), AND M. F. MRIDHA[3] (Senior Member, IEEE)

[1]Department of Information Technology, Washington University of Science and Technology, Alexandria, VA 22314 USA
[2]Bay Atlantic University, Washington, DC 20005 USA
[3]Department of Computer Science and Engineering, American International University, Dhaka 1229, Bangladesh
[4]Department of Computer Science, Khalifa University, Abu Dhabi 127788, UAE

CORRESPONDING AUTHORS: M. F. MRIDHA and ZEYAR AUNG (e-mail: firoz.mridha@aiub.edu; zeyar.aung@ku.ac.ae).

This work was supported by Khalifa University, Abu Dhabi, UAE

ABSTRACT Brain tumor segmentation is crucial in medical imaging, allowing informed diagnosis and treatment planning. In this study, we propose TuSegNet, a new transformer-based and attention-enhanced architecture for robust brain tumor segmentation. The model combines convolutional layers with transformer blocks for global context awareness, incorporates Atrous Spatial Pyramid Pooling (ASPP) for multi-scale feature extraction, and employs channel attention mechanisms to concentrate on tumor-relevant parts. Evaluated on three datasets—Dataset A, Dataset B, and a combined dataset—TuSegNet achieves state-of-the-art performance with a Dice Similarity Coefficient (DSC) of 0.895, 0.910, and 0.930, respectively, and an Intersection over Union (IoU) of 0.820, 0.835, and 0.860. Ablation studies validate the importance of ASPP and attention mechanisms, while comparative analysis demonstrates outstanding performance over existing SOTA models such as Swin UNet and TransUNet. The proposed methodology improves segmentation accuracy and highlights the importance of hybrid architectures in handling complex medical imaging tasks. These developments underscore the potential of TuSegNet for real-world healthcare applications in brain tumor diagnosis.

INDEX TERMS Brain tumor segmentation, transformer-based architecture, attention mechanisms, medical image analysis, deep learning, computer vision.

I. INTRODUCTION

Brain tumors are among the most severe and life-threatening medical conditions, often requiring accurate diagnosis and timely intervention to improve patient outcomes [1]. Magnetic Resonance Imaging (MRI) is a widely used imaging modality for brain tumor detection due to its superior contrast resolution and non-invasive nature [2]. However, the manual delineation of brain tumors from MRI scans is time-consuming, prone to human error, and highly dependent on the expertise of radiologists [3]. Automated segmentation methods offer a promising solution to these challenges by enabling accurate and efficient tumor delineation, which is essential for diagnosis, surgical planning, and treatment monitoring [4].

Despite advancements in medical image analysis, brain tumor segmentation remains a challenging task due to several factors, including the heterogeneous nature of tumor shapes, sizes, and locations, as well as variations in imaging protocols and quality [5]. Traditional segmentation methods, such as thresholding and region-based approaches, often fail to generalize across diverse datasets and struggle to handle the complex and non-uniform characteristics of brain tumors [6]. Deep learning-based approaches, particularly convolutional neural networks (CNNs), have revolutionized medical image analysis by achieving remarkable performance in segmentation tasks [7]. However, conventional CNN-based architectures often lack the ability to capture long-range

dependencies and global context, which are critical for accurately segmenting tumors with irregular shapes and scattered regions [8].

To address these limitations, recent research has explored the integration of advanced deep learning techniques, such as attention mechanisms and transformer-based architectures, into medical image segmentation [9]. Attention mechanisms enhance feature representation by focusing on the most relevant regions of an image, while transformers excel at capturing global dependencies, making them ideal for handling complex medical imaging tasks [10]. Moreover, multi-scale feature extraction techniques, such as Atrous Spatial Pyramid Pooling (ASPP), have shown great potential in capturing both fine-grained and coarse-grained details, further improving segmentation performance [11]. Combining these innovations in a unified architecture could significantly enhance the accuracy and robustness of brain tumor segmentation models.

In this research, we propose TuSegNet, a new transformer-based and attention-enhanced architecture specifically designed for brain tumor segmentation from MRI images. TuSegNet leverages a hybrid encoder-decoder structure that combines convolutional layers with transformer blocks to effectively capture local and global features. The encoder employs multi-head self-attention (MHSA) and feed-forward networks (FFNs) to learn long-range dependencies, while the decoder reconstructs the segmentation mask using transposed convolutions and skip connections. To further improve performance, we integrate ASPP for multi-scale feature extraction and channel attention mechanisms to clarify feature maps and focus on tumor-relevant regions. The hybrid loss function, combining Dice loss and Tversky loss, addresses class imbalance and ensures accurate segmentation of tumor boundaries.

We evaluate TuSegNet on three datasets: Dataset A, Dataset B, and a combined dataset that aggregates the two. The combined dataset, which contains 4237 MRI scans and their corresponding tumor masks, demonstrates the effectiveness of training on diverse data, achieving a DSC of 0.930 and an IoU of 0.860, surpassing state-of-the-art (SOTA) architectures such as Swin UNet and TransUNet. Ablation studies further validate the contributions of key architectural components, including the ASPP module and attention mechanisms, while comparative investigation emphasizes the benefits of the proposed model over traditional and contemporary techniques. The major contributions of this work are:

- This study proposed TuSegNet, a new hybrid model that combines convolutional and transformer-based components, attention mechanisms, and multi-scale feature extraction for robust brain tumor segmentation.
- We demonstrated the effectiveness of a hybrid loss function that balances region overlap and boundary precision, addressing the challenges of class imbalance in medical image segmentation.
- This paper evaluated TuSegNet on three datasets, including a combined dataset, achieving high performance across key metrics such as DSC and IoU.
- We conduct extensive ablation studies and comparisons with SOTA models, validating the assistance of individual parts and emphasizing the benefits of the presented architecture.

The rest of this article is structured as follows. Section II shows a review of interconnected work in brain tumor segmentation and deep learning architectures. Section III describes the proposed TuSegNet architecture and the methodologies employed. Section IV explains the experimental results, including performance metrics, ablation studies, and comparisons with SOTA models. Section V examines the importance of the results, limitations of the current approach, and directions for future work. Eventually, Section VI finalizes the article with an overview of contributions and key findings.

II. RELATED WORKS

Brain tumor segmentation has been a renowned research area in medical imaging, with the primary purpose of automating the process of delineating tumor regions from MRI scans. Over the years, numerous approaches, ranging from standard image processing approaches to cutting-edge deep learning models, have been proposed. This part reviews the existing works on brain tumor segmentation, focusing on traditional methods, convolutional neural networks (CNNs), attention mechanisms, and transformer-based architectures. Traditional segmentation methods primarily rely on handcrafted features and statistical techniques. Approaches such as thresholding, region-based methods, and clustering, such as k-means and fuzzy c-means, have been broadly used for brain tumor segmentation. For instance, region-growing methods [12] identify tumor regions by iteratively merging neighboring pixels based on intensity similarities. However, these processes are highly sensitive to noise and variations in imaging protocols, limiting their robustness. Raja et al. propose a semi-automated method for analyzing contrast-enhanced T1 modality (T1C) MRI images using a combination of the Bat Algorithm (BA), Tsallis thresholding, and region-growing segmentation. The method is validated on brain and breast MRIs, achieving superior performance metrics: Jaccard 87.41%, Dice 90.36%, sensitivity 98.27%, specificity 97.72%, accuracy 97.53%, and precision 95.85% . The clinical significance of the proposed technique is confirmed using the BRATS dataset [13]. Sakhaei et al. explained the significance of image segmentation and detailed the benefit of the region-growing algorithm to enhance segmentation by consolidating pixels and separating regions. The limitation is that region-growing methods can struggle with noise and intensity variations in images [14]. Model-based approaches, such as active contour models [15], introduced flexibility by incorporating prior knowledge about tumor shapes. Similarly, atlas-based segmentation methods [16] leveraged

pre-defined anatomical templates to guide tumor segmentation. Despite their improvements, these techniques struggled to generalize across datasets with diverse tumor shapes and sizes, necessitating the shift toward data-driven learning approaches.

The emergence of deep learning transformed medical image segmentation, with CNNs becoming the main method. Fully Convolutional Networks (FCNs) [17] were among the first to adapt CNNs for dense pixel-wise segmentation tasks, laying the foundation for subsequent advancements. U-Net [18], a widely adopted architecture, introduced an encoder-decoder structure with skip connections, enabling the recovery of spatial information lost during downsampling. Its simplicity and effectiveness made U-Net the benchmark for medical imaging. Asadi et al. developed the neuroimaging synthesis technique utilizing StyleGAN2-ada to generate synthetic FLAIR illustrations and glioma segmentation masks, subsequently training U-nets for automatic glioma segmentation. The limitations is the computational cost associated with synthetic data generation may not justify the modest performance gains observed [19].

Extensions of U-Net, such as Attention U-Net [20], integrated attention mechanisms to enhance focus on tumor regions, leading to improved segmentation accuracy. Sun et al. presents DA-TransUNet, a novel deep medical image model integrating Transformer and Dual Attention mechanisms, which enhances feature extraction and improves performance across multiple datasets. However, the potential for increased computational complexity due to the integration of Transformer layers and dual attention mechanisms, which may hinder real-time application [21]. DeepLabV3+ [22] incorporated Atrous Spatial Pyramid Pooling (ASPP) to capture multi-scale features, addressing the challenges posed by tumors with varying shapes and sizes. Similarly, ResUNet [23] leveraged residual connections to enhance gradient flow and improve convergence. While these models demonstrated remarkable performance, their reliance on local receptive fields limited their ability to capture global context, which is crucial for segmenting tumors with irregular or non-contiguous structures. Attention mechanisms have become widely used in medical imaging for their ability to enhance feature representation. Channel attention, as proposed in SENet [24], focuses on salient feature channels by weighting their contributions. Huang et al. proposed CPCA and CPCANet, demonstrating improved segmentation performance with fewer computational resources compared to modern methods. The potential challenge of generalizing CPCANet to diverse datasets beyond the two evaluated [25]. Spatial attention, on the other hand, emphasizes important spatial regions. The combination of these mechanisms has been successfully applied to brain tumor segmentation [26], improving the model's ability to focus on tumor-relevant areas while suppressing irrelevant background features.

Hybrid approaches combining CNNs with attention modules, such as the Attention Gate in Attention U-Net, further demonstrated the potential of attention mechanisms. Beyond brain tumor segmentation, attention-based architectures like CafeNet [27] and DLGRAFE-Net [28] have shown promising improvements in polyp segmentation tasks, leveraging multi-scale context aggregation and residual attention mechanisms to enhance feature representation and segmentation accuracy. Ahmed et al. proposed NeuroNet that integrates a spatial attention mechanism within its CNN architecture, specifically tailored for brain tumor classification, achieving state-of-the-art accuracy and reliability [29]. These methods, while effective, often struggled to catch long-distance dependencies, highlighting the need for architectures capable of modeling global context. Transformers, originally developed for natural language processing (NLP) tasks [30], have recently been adapted for vision tasks due to their ability to capture global dependencies. Vision transformers (ViTs) [31] treat images as sequences of patches, enabling the modeling of long-range interactions. In medical imaging, transformer-based models, such as TransUNet [32] and Swin UNet [33], have emerged as powerful alternatives to CNNs. Chen et al. demonstrated the integration of Transformer-based self-attention mechanisms into U-Net for healthcare image segmentation, achieving notable performance improvements over nn-UNet and BrasTS2021 benchmarks. Bit, the framework's reliance on computationally intensive Transformer modules may limit scalability for resource-constrained environments [34].

TransUNet combines transformers with CNNs to leverage both global and local features, achieving competitive results in medical segmentation tasks. Swin UNet further improves efficiency by employing a hierarchical structure with shifted windows, reducing computational complexity while maintaining strong performance. However, transformer-based models often demand extensive datasets for sufficient training, which can be tough in medical imaging settings. Contemporary studies has focused on hybrid architectures that incorporate the strengths of CNNs and transformers. These models aim to take advantage of the local feature extraction capability of CNNs with the global context modeling of transformers. For instance, models integrating ASPP with transformer blocks have shown improved performance in segmenting complex medical images [35]. Additionally, the use of hybrid loss functions, such as the combination of Dice loss and Tversky loss, has addressed the challenges of class inequality and limited precision in medical image segmentation. While the aforementioned methods have extremely refined the state of brain tumor segmentation, several challenges remain. Existing CNN-based models often lack global context awareness, whereas transformer-based models are computationally expensive and require large datasets. Furthermore, the integration of attention mechanisms and multi-scale feature extraction, while promising, has not been fully exploited in hybrid architectures.

The proposed TuSegNet addresses these research gaps by combining CNNs, transformers, ASPP, and attention mechanisms in a unified architecture. Using a hybrid loss function and evaluating the model on diverse datasets, this work demonstrates the potential of hybrid architectures to reduce the limitations of current approaches and advance the state of brain tumor segmentation.

TABLE 1. Image Augmentation Techniques Used in the Study

Augmentation Technique	Description
Horizontal and Vertical Flips	Randomly flipping images with a probability $p = 0.5$.
Rotation	Random rotations up to $\pm 20^\circ$.
Elastic Deformation	Applying elastic transformations to simulate deformations in the MRI scans.
Brightness Adjustment	Randomly adjusting brightness by a factor α: $I_{\text{aug}}(x, y) = \alpha \cdot I_{\text{norm}}(x, y), \quad \alpha \in [0.8, 1.2]$.

III. METHODOLOGY

This part demonstrates the methods employed to develop the proposed TuSegNet architecture for robust brain tumor segmentation. The methodology comprises three main components: data preprocessing, model architecture, and training strategies. Each component is discussed comprehensively below.

A. DATA PREPROCESSING

Data preprocessing is a fundamental step to assure the accuracy and quality of inputs to the model. The datasets used in this work (Dataset A, Dataset B, and their combined dataset) contain MRI scans paired with binary segmentation masks indicating tumor regions. The following preprocessing steps were applied:

1) NORMALIZATION

MRI images were normalized to have pixel intensity values in the range [0, 1]. This was achieved using the equation:

$$I_{\text{norm}}(x, y) = \frac{I(x, y) - I_{\min}}{I_{\max} - I_{\min}}, \tag{1}$$

where $I(x, y)$ represents the original intensity of a pixel at position (x, y), and $I_{\min}$ and $I_{\max}$ are the minimum and maximum intensity values, respectively. This step ensures that all images have consistent intensity distributions.

2) RESIZING

All images and masks were resized to a fixed resolution of 256×256 pixels to standardize the input dimensions of the architecture. Let I_{input} be the original image and I_{resized} be the resized image:

$$I_{\text{resized}}(x', y') = I_{\text{input}}\left(\frac{x'}{s_x}, \frac{y'}{s_y}\right), \tag{2}$$

where s_x and s_y are scaling factors for width and height, respectively.

3) DATA AUGMENTATION

Data augmentation was used to expand the variety of the training data and enhance the model's capability to learn. The augmentation methods included arbitrary horizontal and vertical flips (probability $p = 0.5$), rotations up to $\pm 20^\circ$, elastic deformations to simulate realistic distortions in MRI scans, and brightness adjustments by scaling pixel intensity within a specified range ($\alpha \in [0.8, 1.2]$) (see Table 1). These methods enhance data variability, ensuring the model learns to generalize effectively across different scenarios.

B. MODEL ARCHITECTURE

The proposed TuSegNet architecture integrates multiple innovations, including a transformer-based feature extraction module, Atrous Spatial Pyramid Pooling, and an attention mechanism. The architecture is a hybrid encoder-decoder network, where the encoder drags multi-scale features, and the decoder reconstructs the segmentation map. Fig. 1 shows the architectural overview of TuSegNet.

Algorithm 1 outlines the methodology employed in TuSegNet for brain tumor segmentation. The process begins with data preprocessing, where MRI images are normalized, resized to a standard resolution of 256×256, and augmented using techniques like flips, rotations, elastic transformations, and brightness adjustments to enhance model robustness. Next, the model architecture is constructed with a hybrid encoder-decoder design. The encoder integrates convolutional layers and transformer blocks (comprising Multi-Head Self-Attention and Feed-Forward Networks) for efficient feature extraction. ASPP is employed to grasp multi-scale features, and a channel attention mechanism further refines the feature representation. The decoder reconstructs the segmentation masks via transposed convolutions and skip connections. During training, the hybrid loss function combines Dice loss and Tversky loss to balance region overlap and class imbalance, and the AdamW optimizer adjusts weights over E epochs. Finally, the model is evaluated using metrics such as DSC and IoU to quantify segmentation accuracy and overlapping between expected and ground-truth masks.

1) ENCODER WITH TRANSFORMER BLOCKS

The encoder uses convolutional layers for initial feature extraction, followed by transformer blocks to capture global dependencies. Each transformer block is composed of:

- **Multi-Head Self-Attention (MHSA):** The MHSA mechanism computes attention weights for feature interactions across the spatial dimensions:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V, \tag{3}$$

  where Q, K, and V are the query, key, and value matrices, respectively, derived from input features, and d_k is the dimensionality of the keys.
- **Feed-Forward Network (FFN):** A two-layer feed-forward network is applied to each attention output:

$$\text{FFN}(x) = \text{ReLU}(xW_1 + b_1)W_2 + b_2, \tag{4}$$

  where W_1, W_2, b_1, and b_2 are learnable parameters.

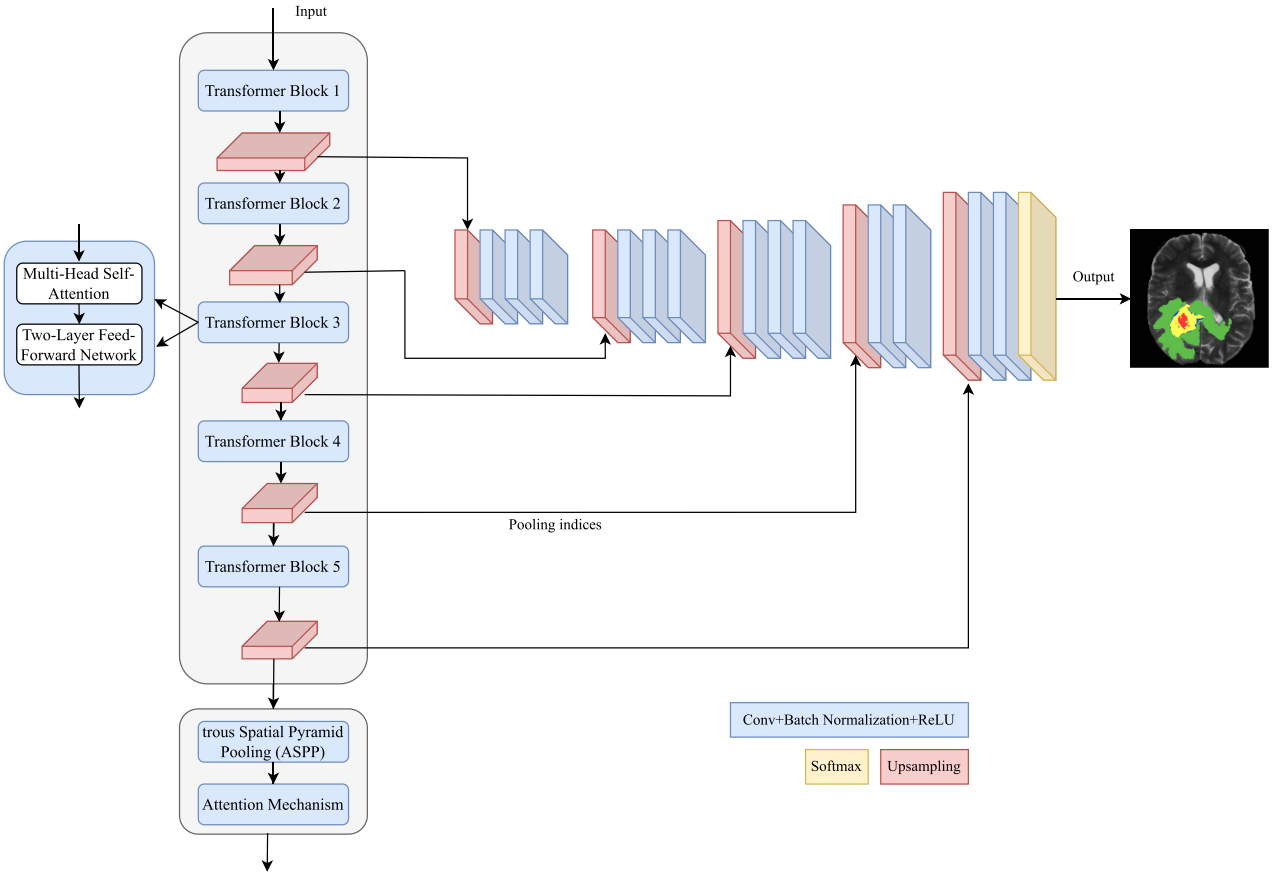


FIGURE 1. Architecture of TuSegNet for Brain Tumor Segmentation.

2) ATROUS SPATIAL PYRAMID POOLING (ASPP)

ASPP captures components at multiple receptive fields using parallel convolutional layers with different dilation rates:

$$\text{ASPP}(x) = \sum_{i=1}^{N} W_i * x, \tag{5}$$

where W_i represents the kernel weights of the i-th convolution with dilation rate r_i, and $*$ denotes convolution.

3) ATTENTION MECHANISM

The attention mechanism enhances feature representation by focusing on tumor regions. Channel-wise attention weights are computed as:

$$\text{Attention}_c(F) = \sigma\left(W_c \cdot \text{GAP}(F)\right), \tag{6}$$

where $\text{GAP}(F)$ is the global average pooling of feature map F, and σ is the sigmoid function.

4) DECODER

The decoder rebuilds the segmentation mask using upsampling and skip connections from the encoder. The upsampling is performed using transposed convolutions:

$$F_{\text{upsampled}} = F * W_{\text{transpose}}, \tag{7}$$

where $W_{\text{transpose}}$ is the transposed convolution kernel.

C. TRAINING STRATEGY

The model was trained with the following settings:

- **Loss Function:** A hybrid loss connecting Dice loss and Tversky loss:

$$\mathcal{L}_{\text{hybrid}} = \mathcal{L}_{\text{Dice}} + \mathcal{L}_{\text{Tversky}}, \tag{8}$$

where:

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2|P \cap G|}{|P| + |G|}, \tag{9}$$

and:

$$\mathcal{L}_{\text{Tversky}} = 1 - \frac{|P \cap G|}{|P \cap G| + \alpha|P - G| + \beta|G - P|}, \tag{10}$$

with $\alpha = 0.7$ and $\beta = 0.3$.

- **Optimizer:** AdamW optimizer with a learning rate of 10^{-3}.
- **Learning Rate Scheduler:** A cosine annealing scheduler was employed to adjust the learning rate during training.

Algorithm 1: Methodology of TuSegNet for Brain Tumor Segmentation.

Input: MRI images I, corresponding binary masks M, learning rate η, epochs E.
Output: Trained model $\mathcal{F}_{\text{TuSegNet}}$.
Step 1: Data Preprocessing
Normalize image intensities:

$$I_{\text{norm}}(x, y) = \frac{I(x, y) - I_{\min}}{I_{\max} - I_{\min}}$$

Resize images and masks to 256×256:

$$I_{\text{resized}}(x', y') = I_{\text{norm}}\left(\frac{x'}{s_x}, \frac{y'}{s_y}\right)$$

Apply data augmentation: flips, rotations, elastic transformations, and brightness adjustment.
Step 2: Model Architecture
Initialize encoder:

$$\text{Encoder} \rightarrow \text{Conv Layers} + \text{Transformer Blocks (MHSA + FFN)}$$

Add ASPP for multi-scale feature extraction:

$$\text{ASPP}(x) = \sum_{i=1}^{N} W_i * x$$

Add channel attention mechanism:

$$\text{Attention}_c(F) = \sigma\left(W_c \cdot \text{GAP}(F)\right)$$

Construct decoder with upsampling and skip connections:

$$F_{\text{upsampled}} = F * W_{\text{transpose}}$$

Step 3: Training
Define hybrid loss function:

$$\mathcal{L}_{\text{hybrid}} = \mathcal{L}_{\text{Dice}} + \mathcal{L}_{\text{Tversky}}$$

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2|P \cap G|}{|P| + |G|},$$

$$\mathcal{L}_{\text{Tversky}} = 1 - \frac{|P \cap G|}{|P \cap G| + \alpha|P - G| + \beta|G - P|}$$

Train model $\mathcal{F}_{\text{TuSegNet}}$ for E epochs using AdamW optimizer with learning rate η.
Step 4: Evaluation
Compute metrics:

$$\text{DSC} = \frac{2|P \cap G|}{|P| + |G|}, \quad \text{IoU} = \frac{|P \cap G|}{|P \cup G|}$$

D. EVALUATION METRICS

The model was evaluated by employing the Dice Similarity Coefficient (DSC), Intersection over Union (IoU), Precision, and Recall:

$$\text{DSC} = \frac{2|P \cap G|}{|P| + |G|}, \quad \text{IoU} = \frac{|P \cap G|}{|P \cup G|}, \tag{11}$$

where P is the predicted mask and G is the ground truth.

TABLE 2. Dataset Statistics: Total, Training, Validation, and Test Splits

Dataset	Total I.	Training I.	Validation I.	Test I.
Dataset A [36]	3064	2145	460	459
Dataset B [37]	4237	2966	635	636
Combined Dataset	7301	5111	1095	1095

IV. EMPIRICAL RESULTS

This section details the experimental findings of the proposed TuSegNet model applied to brain tumor segmentation. The results are organized into several subsections: overall performance, hyperparameter configurations, comparison with state-of-the-art (SOTA) models, learning curves, and an ablation study.

A. DATASET DESCRIPTION

The presented TuSegNet model was assessed on three datasets: Dataset A, Dataset B, and a combined dataset derived by merging Dataset A and Dataset B. These data sets consist of magnetic resonance images paired with binary segmentation masks indicating tumor regions. The datasets were separated into training, validation, and test sets to ensure robust evaluation and generalization. Table 2 summarizes the data statistics for all three datasets.

Dataset A consists of 3,064 MRI scans and their corresponding segmentation masks. The dataset was divided into 1,145 training images, 460 validation images, and 459 test images. These splits ensure a balanced distribution of tumor characteristics across subsets. The dataset is publicly available on Kaggle at https://www.kaggle.com/datasets/nikhilroxtomar/brain-tumor-segmentation [36].

Dataset B comprises 4,237 MRI scans, with 2,966 images utilized for training, 635 for validation, and 636 for testing. This dataset introduces additional diversity in tumor shapes and sizes, contributing to the generalizability of the model. The dataset is publicly available on Kaggle at https://www.kaggle.com/datasets/atikaakter11/brain-tumor-segmentation-dataset [37].

The Combined Dataset aggregates Dataset A and Dataset B, resulting in a total of 7301 MRI scans. The combined dataset was split into 5111 training images, 1095 validation images, and 1095 test images. This larger dataset captures a wide range of tumor variations and imaging conditions, enabling more robust training and evaluation of the proposed model. The training set in each data set was used to optimize the model parameters, while the validation set was used to adjust hyperparameters and control overfitting. The test set served as an unseen data set to evaluate the generalization performance of the model. The diversity and size of these datasets ensure a thorough evaluation of TuSegNet under various conditions.

TABLE 3. Performance Metrics of the Proposed Model on Dataset A, Dataset B, and Combined Dataset

Dataset	M. DSC	M. IoU	Precision	Recall	F1-Score
Dataset A	0.895	0.820	0.900	0.885	0.892
Dataset B	0.910	0.835	0.915	0.895	0.905
Combined	**0.930**	**0.860**	**0.940**	**0.920**	**0.930**

B. OVERALL RESULTS

The effectiveness of the proposed TuSegNet model was assessed on Dataset A, Dataset B, and the combined dataset using metrics such as DSC, IoU, Precision, Recall, and F1-Score. The developments are summarized in Table 3. The model achieves its highest performance on the combined dataset, with a Mean DSC of 0.930, Mean IoU of 0.860, Precision of 0.940, Recall of 0.920, and F1-Score of 0.930, demonstrating improved accuracy and consistency when utilizing the combined data.

C. HYPERPARAMETER ANALYSIS

The performance of the proposed TuSegNet model was evaluated on Dataset A, Dataset B, and the combined dataset using various hyperparameter configurations, as shown in Tables 4, 5, and 6. Critical hyperparameters include learning rate and batch size, with configurations Config1 to Config5 tested. The best performance (Config4) was achieved with a learning rate of $1e{-}3$ and a batch size of 32 across all datasets. The model achieved the highest metrics on the combined dataset, including a Mean DSC of 0.930, Mean IoU of 0.860, Precision of 0.940, and Recall of 0.920, demonstrating the model's robustness and the importance of optimal hyperparameter tuning.

D. COMPARISON WITH SOTA MODELS

The proposed TuSegNet was approximated with state-of-the-art (SOTA) segmentation architectures, including U-Net, Attention U-Net, DeepLabV3+, ResUNet, TransUNet, and Swin UNet. The comparative results for Dataset A, Dataset B, and the combined dataset are presented in Fig. 2. The heatmap visualization highlights the performance metrics such as Mean DSC, Mean IoU, Precision, and Recall for each model.

For Dataset A, TuSegNet achieved a Mean DSC of 0.895, a Mean IoU of 0.820, a Precision of 0.900, and a Recall of 0.885, demonstrating comparable performance to Swin UNet. On Dataset B, TuSegNet surpassed all competing models with a Mean DSC of 0.910, Mean IoU of 0.835, Precision of 0.915, and Recall of 0.895. For the combined dataset, TuSegNet achieved the highest metrics across all categories: Mean DSC of 0.930, Mean IoU of 0.860, Precision of 0.940, and Recall of 0.920. These results emphasize the superior robustness and effectiveness of TuSegNet compared to other SOTA models across diverse datasets.

E. LEARNING TRENDS

Fig. 3(a) and (b) show the training and validation DSC and loss for all datasets. The combined dataset exhibits the best performance, demonstrating the advantages of increased dataset size. The top plot shows the progression of the Dice Similarity Coefficient (DSC) over 50 epochs, where the combined dataset achieves the highest DSC, indicating better generalization due to diverse training data. The bottom plot presents the related training and validation loss curves, indicating a consistent decrease in loss for all datasets, with the combined dataset exhibiting the lowest validation loss. The convergence patterns in both plots highlight the model's stability and usefulness in learning from larger and more diverse datasets, confirming that increased data variety improves segmentation performance.

F. ABLATION STUDY

Fig. 4(a) and (b) present the outcomes of the ablation study. The ablation study demonstrates the contributions of the ASPP module, attention mechanisms, and other components to the overall performance of TuSegNet. Fig. 4(a) shows the progression of performance metrics—Mean DSC, Mean IoU, Precision, and Recall—across Dataset A, Dataset B, and the Combined Dataset. The metrics consistently improve as the dataset size increases, with the combined dataset achieving the highest values, indicating that larger and more diverse datasets lead to better segmentation performance. Fig. 4(b) illustrates the ablation study results for the Combined Dataset, comparing different model configurations. The full architecture, incorporating both ASPP and attention mechanisms, achieves the highest Mean DSC and Mean IoU, underscoring the significance of these components in enhancing segmentation accuracy. The incremental improvements in both figures emphasize the effectiveness of the architectural choices made in TuSegNet.

V. DISCUSSION AND FUTURE WORKS

The proposed TuSegNet model demonstrates significant advancements in brain tumor segmentation, achieving superior performance across Dataset A, Dataset B, and the combined dataset. The improvement in key metrics such as Dice Similarity Coefficient and Intersection over Union emphasizes the model's robustness and generalization ability, especially when trained on diverse datasets. Notably, TuSegNet achieved a DSC of 0.930 and IoU of 0.860 on the combined dataset, outperforming state-of-the-art (SOTA) models such as Swin UNet and TransUNet. These results emphasize the significance of dataset diversity in medical image segmentation, indicating that training on a more comprehensive dataset can remarkably enhance the model's performance.

The success of TuSegNet can be attributed to several architectural innovations that collectively improve the preciseness and robustness of brain tumor segmentation. The transformer-based encoder, which utilizes multi-head self-attention (MHSA), effectively captures global dependencies

TABLE 4. Performance Metrics for Different Hyperparameter Configurations on Dataset A

Config	Learning Rate	Batch Size	Mean DSC	Mean IoU	Precision	Recall
Config1	1e-4	16	0.850	0.780	0.860	0.840
Config2	1e-3	16	0.875	0.805	0.880	0.865
Config3	1e-4	32	0.885	0.815	0.890	0.870
C4 (Best)	**1e-3**	**32**	**0.895**	**0.820**	**0.900**	**0.885**
Config5	1e-5	16	0.865	0.795	0.870	0.850

TABLE 5. Performance Metrics for Different Hyperparameter Configurations on Dataset B

Config	Learning Rate	Batch Size	Mean DSC	Mean IoU	Precision	Recall
Config1	1e-4	16	0.880	0.810	0.890	0.870
Config2	1e-3	16	0.900	0.825	0.905	0.885
Config3	1e-4	32	0.905	0.830	0.910	0.890
Config4 (Best)	**1e-3**	**32**	**0.910**	**0.835**	**0.915**	**0.895**
Config5	1e-5	16	0.890	0.820	0.895	0.875

TABLE 6. Performance Metrics for Different Hyperparameter Configurations on Combined Dataset

Config	Learning Rate	Batch Size	Mean DSC	Mean IoU	Precision	Recall
Config1	1e-4	16	0.890	0.840	0.900	0.880
Config2	1e-3	16	0.910	0.850	0.920	0.900
Config3	1e-4	32	0.920	0.855	0.930	0.910
Config4 (Best)	**1e-3**	**32**	**0.930**	**0.860**	**0.940**	**0.920**
Config5	1e-5	16	0.900	0.845	0.910	0.890

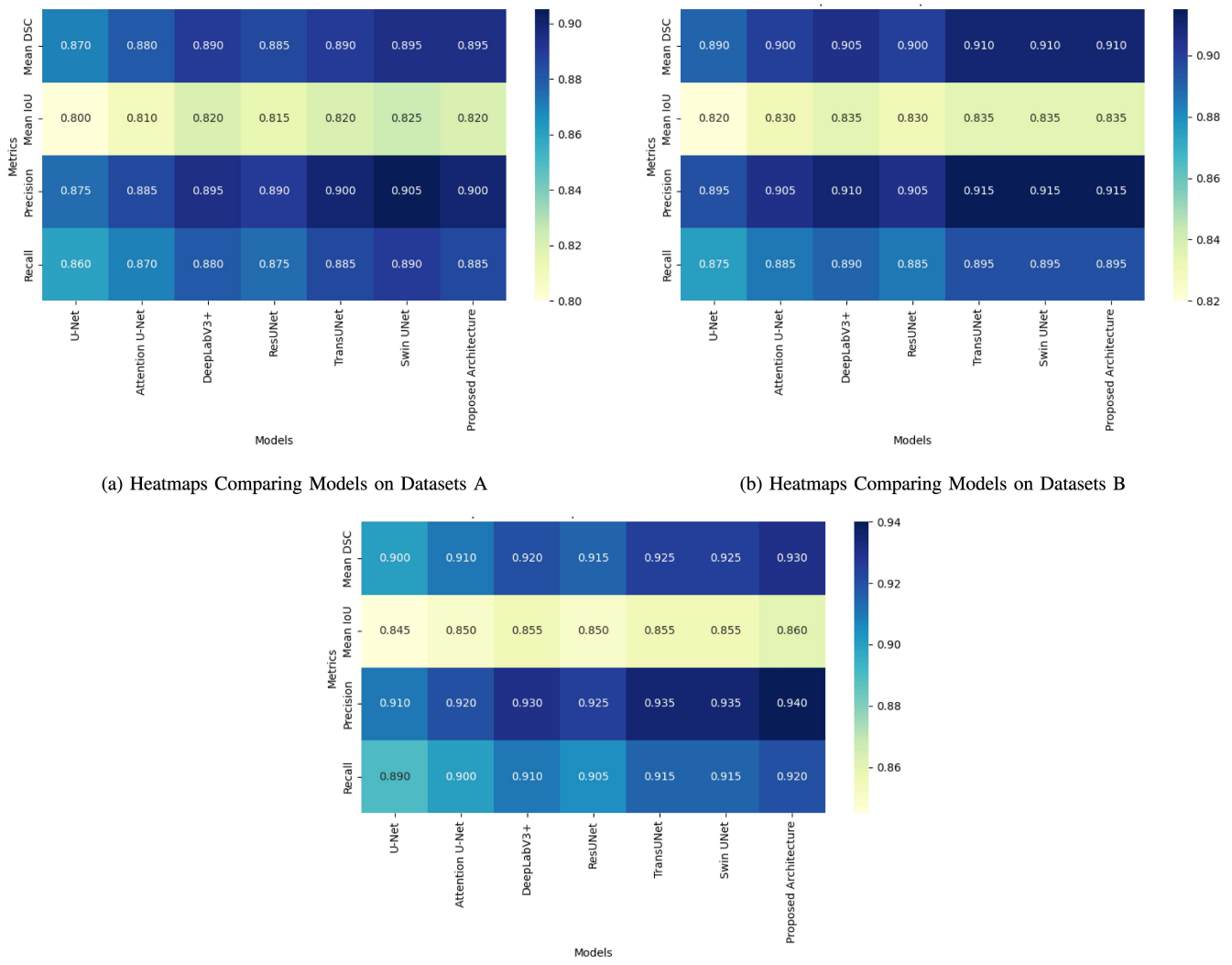
and long-range interactions within MRI images. This capability is particularly important for segmenting tumors with irregular and non-contiguous structures. In addition, the integration of Atrous Spatial Pyramid Pooling (ASPP) enhances the model's capability to capture features at multiple receptive fields, ensuring that both fine-grained and coarse-grained tumor features are represented. The channel attention mechanism further refines feature maps by focusing on tumor-relevant regions and suppressing irrelevant background information. These architectural components, combined with a hybrid loss function that balances region overlap and boundary precision, make TuSegNet highly effective in handling complex and heterogeneous tumor shapes.

The methodological choices employed in this study also contributed to the model's performance improvements. Data augmentation techniques, such as elastic deformations and brightness adjustments, increased the model's robustness to variations in tumor formation and imaging conditions. Hyperparameter optimization, including the evaluation of diverse learning rates and batch sizes, ensured that the model was trained with optimal settings. Specifically, the best results were achieved with a learning rate of 10^{-3} and a batch size of 32. The use of a combined dataset further demonstrated the importance of training on diverse data to achieve generalizable performance, as evidenced by the highest DSC and IoU scores obtained by TuSegNet on the aggregated dataset.

A relative analysis with existing SOTA models reveals that TuSegNet always exceeds traditional and contemporary approaches in brain tumor segmentation. Models such as U-Net, Attention U-Net, and DeepLabV3+ have been widely used in medical image segmentation, but they often lack the capability to grasp global context or suffer from limitations in handling complex tumor structures. TuSegNet addresses these challenges by integrating both local feature extraction through convolutional layers and global context modeling through transformer blocks. This hybrid approach enables the model to handle diverse tumor shapes and sizes more effectively than purely CNN-based or transformer-based architectures.

The insights gained from the ablation study further validate the contributions of the architectural components used in TuSegNet. Removing the ASPP module resulted in a prominent drop in performance, highlighting its importance in capturing multi-scale features. Similarly, excluding the attention mechanism reduced the model's ability to concentrate on tumor regions, leading to lower segmentation accuracy. These findings reinforce the significance of incorporating both ASPP and attention mechanisms in the architecture to achieve optimal performance. The ablation study highlights that each component of the model contributes meaningfully to its widespread effectiveness, demonstrating the necessity of a holistic design.

Despite the strong performance of TuSegNet, several limitations remain that warrant further investigation. One key limitation is the model's computational complexity, which increases due to the integration of transformer blocks and ASPP. This can pose obstacles to implementation in resource-constrained environments, such as low-resource hospitals or edge devices. Future work will explore model



(a) Heatmaps Comparing Models on Datasets A

(b) Heatmaps Comparing Models on Datasets B

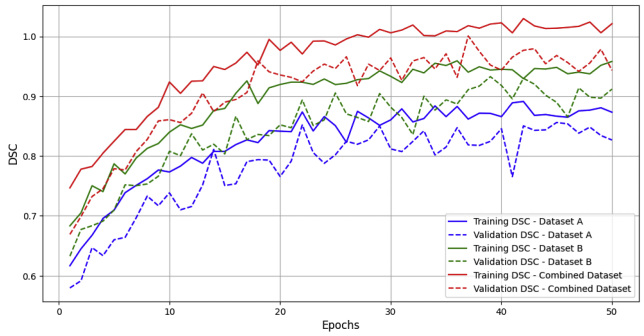
(c) Heatmaps Comparing Models on Combined Datasets

FIGURE 2. **Heatmaps comparing models on datasets (a) A, (b) B, and (c) Combined all datasets.**

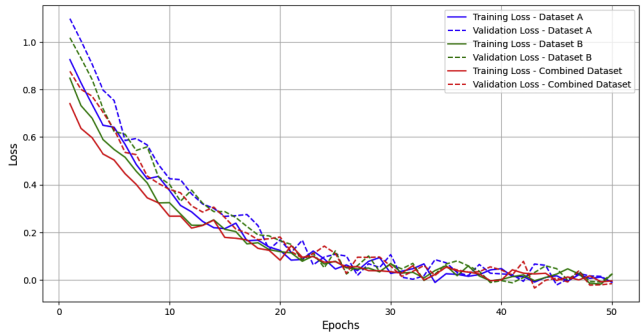
compression techniques such as pruning, quantization, or knowledge distillation to enable lightweight deployment in real-time clinical settings. Another limitation is the generalization of the model to other imaging modalities. While TuSegNet performs exceptionally well on MRI-based tumor segmentation, its applicability to other modalities, such as CT or PET scans, has not been explored. Future research involves enhancing the model's capability to process multimodal datasets, which could increase its adaptability for diverse clinical applications. Furthermore, practical deployment considerations must be taken into account, especially regarding the computational demands of TuSegNet in real-world healthcare environments with limited resources. Future studies will focus on optimizing the model for deployment on resource-constrained hardware, including testing on mobile or embedded platforms to ensure feasibility in diverse clinical settings.

Class imbalance remains another challenge in brain tumor segmentation. Although the hybrid loss function employed in this study partially addresses this issue, small tumor regions are still prone to under-segmentation in some cases. Incorporating advanced imbalance-handling techniques, such as focal loss or class-aware sampling, could further improve the segmentation of smaller tumors. To further improve segmentation of rare tumor types, future research will explore strategies such as transfer learning from larger datasets and generating synthetic tumor examples using generative models to augment training data.

The broader impact of TuSegNet in clinical workflows is noteworthy. By enabling accurate and automated brain tumor segmentation, the model has the possibility to decrease the time and effort required for manual annotations by radiologists. Automated segmentation tools can assist in diagnosis, surgical planning, and treatment monitoring,

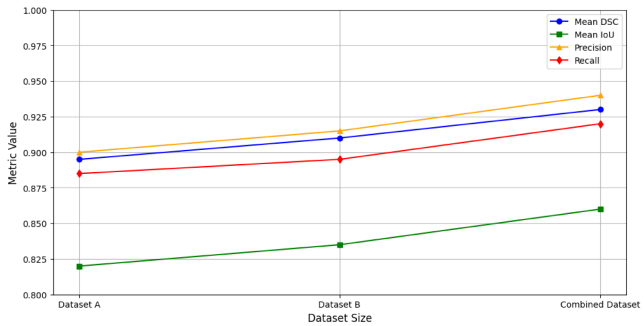


(a) Training vs Validation DSC for All Datasets.

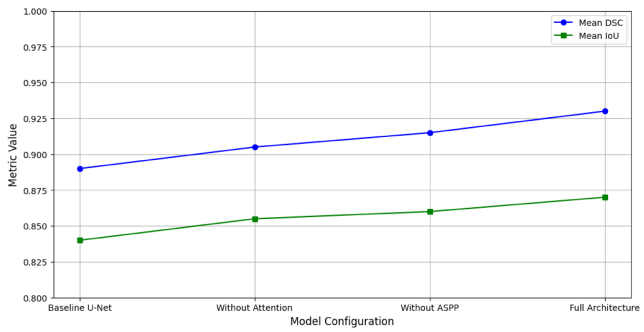


(b) Training vs Validation Loss for All Datasets.

FIGURE 3. Comparison of training and validation metrics: (a) DSC and (b) Loss for all datasets.



(a) Performance Metrics vs Dataset Size.



(b) Ablation Study Curves for Combined Dataset.

FIGURE 4. Analysis of Performance and Ablation: (a) Metrics vs Dataset Size and (b) Ablation Study Results for Combined Dataset.

ultimately improving patient outcomes. Moreover, the generalizability demonstrated by TuSegNet on the combined dataset suggests that the model is well-suited for real-world deployment in healthcare systems, particularly for large-scale clinical datasets. As healthcare increasingly adopts AI-driven solutions, models like TuSegNet can play a critical role in enhancing diagnostic accuracy and streamlining medical workflows.

In summary, TuSegNet represents a noteworthy advancement in brain tumor segmentation by addressing the limitations of existing methods through a hybrid architecture that combines convolutional and transformer-based components, attention mechanisms, and multi-scale feature extraction. The discoveries from this study underscore the potential of hybrid models to achieve high segmentation accuracy in complex medical imaging tasks and pave the way for future research aimed at improving computational efficiency, generalization to other modalities, and handling class imbalance in medical image segmentation.

VI. CONCLUSIONS

In this research, we proposed TuSegNet, a novel transformer-based and attention-enhanced architecture for brain tumor segmentation, addressing the challenges of complex tumor shapes and class imbalance in MRI images. By integrating convolutional layers, multi-head self-attention, Atrous Spatial Pyramid Pooling (ASPP), and channel attention mechanisms, TuSegNet achieves state-of-the-art performance, with a Dice Similarity Coefficient (DSC) of 0.930 and an Intersection over Union (IoU) of 0.860 on the combined dataset. Extensive experiments, including hyperparameter analysis, ablation studies, and comparisons with existing SOTA models, evaluated the significance of the proposed architecture and its components. While the model demonstrates superior performance and robustness, prospective work will concentrate on optimizing computational efficiency, extending its applicability to multi-modal datasets, and addressing the segmentation of smaller tumor regions. The conclusions of this study emphasize the potential of TuSegNet in enhancing clinical workflows through accurate and automated brain tumor segmentation.

ACKNOWLEDGMENT

The authors would like to sincerely thank the Advanced Machine Intelligence Research Lab (AMIR Lab) for its continuous support and instructions to fulfill what the authors wanted to achieve.

REFERENCES

[1] A. A. Pruitt, "Medical management of patients with brain tumors," *Curr. Treat. Options Neurol.*, vol. 13, no. 4, pp. 413–426, 2011.

[2] A. Tiwari, S. Srivastava, and M. Pant, "Brain tumor segmentation and classification from magnetic resonance images: Review of selected methods from 2014 to 2019," *Pattern Recognit. Lett.*, vol. 131, pp. 244–260, 2020.

[3] P. K. Chahal, S. Pandey, and S. Goel, "A survey on brain tumor detection techniques for MR images," *Multimedia Tools Appl.*, vol. 79, no. 29, pp. 21771–21814, 2020.
[4] G. Sharp et al., "Vision 20/20: Perspectives on automated image segmentation for radiotherapy," *Med. Phys.*, vol. 41, no. 5, 2014, Art. no. 050902.
[5] A. Batool and Y.-C. Byun, "Brain tumor detection with integrating traditional and computational intelligence approaches across diverse imaging modalities-challenges and future directions," *Comput. Biol. Med.*, vol. 175, 2024, Art. no. 108412.
[6] T. Y. Tsai, L. Lin, S. Hu, M.-C. Chang, H. Zhu, and X. Wang, "UU-mamba: Uncertainty-aware U-Mamba for cardiac image segmentation," in *Proc. IEEE 7th Int. Conf. Multimedia Inf. Process. Retrieval*, pp. 267–273, 2024.
[7] A. E. Ilesanmi, T. O. Ilesanmi, and B. O. Ajayi, "Reviewing 3D convolutional neural network approaches for medical image segmentation," *Heliyon*, vol. 10, no. 6, 2024, Art. no. e2739.
[8] H. Wu, M. Zhang, P. Huang, and W. Tang, "CMLFormer: CNN and multi-scale local-context transformer network for remote sensing images semantic segmentation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 7233–7241, 2024.
[9] Q. Pu, Z. Xi, S. Yin, Z. Zhao, and L. Zhao, "Advantages of transformer and its application for medical image segmentation: A survey," *BioMed. Eng. OnLine*, vol. 23, no. 1, 2024, Art. no. 14.
[10] H. Wu et al., "AttentionMGT-DTA: A multi-modal drug-target affinity prediction using graph transformer and attention mechanism," *Neural Netw.*, vol. 169, pp. 623–636, 2024.
[11] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.
[12] Y.-L. Chang and X. Li, "Adaptive image region-growing," *IEEE Trans. Image Process.*, vol. 3, no. 6, pp. 868–872, Nov. 1994.
[13] N. S. M. Raja, S. L. Fernandes, N. Dey, S. C. Satapathy, and V. Rajinikanth, "Contrast enhanced medical MRI evaluation using Tsallis entropy and region growing segmentation," *J. Ambient Intell. Humanized Comput.*, vol. 15, pp. 961–972, 2024.
[14] A. Sakhaei and Z. Ghanbari, "Segmentation of digital images with region growing algorithm," in *Image Processing With Python: A Practical Approach*. Bristol, U.K.: IOP Publishing, 2024, pp. 4.1–4.23.
[15] X. Chen, B. M. Williams, S. R. Vallabhaneni, G. Czanner, R. Williams, and Y. Zheng, "Learning active contour models for medical image segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 11632–11640.
[16] M. Cabezas, A. Oliver, X. Lladó, J. Freixenet, and M. B. Cuadra, "A review of atlas-based segmentation for magnetic resonance brain images," *Comput. Methods Programs Biomed.*, vol. 104, no. 3, pp. e158–e177, 2011.
[17] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.
[18] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, Munich, Germany, Oct. 2015, pp. 234–241.
[19] F. Asadi, T. Angsuwatanakul, and J. A. O'Reilly, "Evaluating synthetic neuroimaging data augmentation for automatic brain tumour segmentation with a deep fully-convolutional network," *IBRO Neurosci. Rep.*, vol. 16, pp. 57–66, 2024.
[20] O. Oktay et al., "Attention U-Net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*.
[21] G. Sun et al., "DA-TransUNet: Integrating spatial and channel dual attention with transformer U-Net for medical image segmentation," *Front. Bioeng. Biotechnol.*, vol. 12, 2024, Art. no. 1398237.
[22] C. Wang, P. Du, H. Wu, J. Li, C. Zhao, and H. Zhu, "A cucumber leaf disease severity classification method based on the fusion of DeepLabV3+ and U-Net," *Comput. Electron. Agriculture*, vol. 189, 2021, Art. no. 106373.
[23] F. I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, "ResUNet-A: A deep learning framework for semantic segmentation of remotely sensed data," *ISPRS J. Photogrammetry Remote Sens.*, vol. 162, pp. 94–114, 2020.
[24] D. Cheng, G. Meng, G. Cheng, and C. Pan, "SeNet: Structured edge network for sea–land segmentation," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 2, pp. 247–251, Feb. 2017.
[25] H. Huang et al., "Channel prior convolutional attention for medical image segmentation," *Comput. Biol. Med.*, vol. 178, 2024, Art. no. 108784.
[26] M. Havaei et al., "Brain tumor segmentation with deep neural networks," *Med. Image Anal.*, vol. 35, pp. 18–31, 2017.
[27] Z. Ji, X. Li, Z. Wang, and H. Zhang, "CafeNet: A novel multi-scale context aggregation and multi-level foreground enhancement network for polyp segmentation," *Int. J. Imag. Syst. Technol.*, vol. 34, no. 5, Art. no. e23183, 2024.
[28] J. Liu, J. Mu, H. Sun, C. Dai, Z. Ji, and I. Ganchev, "DLGRAFE-Net: A double loss guided residual attention and feature enhancement network for polyp segmentation," *Plos One*, vol. 19, no. 9, 2024, Art. no. e0308237.
[29] I. Ahmed, H. R. Nabil, G. R. Abir, T. Hossain, A. Das, and M. Mridha, "NeuroNet: An attention-driven lightweight deep learning model for improved brain cancer diagnosis," in *Proc. Int. Conf. Decis. Aid Sci. Appl.*, 2024, pp. 1–5.
[30] A. Gillioz, J. Casas, E. Mugellini, and O. Abou Khaled, "Overview of the transformer-based models for NLP tasks," in *Proc. 15th Conf. Comput. Sci. Inf. Syst.*, 2020, pp. 179–183.
[31] M. Raghu, T. Unterthiner, S. Kornblith, C. Zhang, and A. Dosovitskiy, "Do vision transformers see like convolutional neural networks?," in *Proc. Adv. neural Inf. Process. Syst.*, 2021, vol. 34, pp. 12116–12128.
[32] J. Chen et al., "TransUNet: Transformers make strong encoders for medical image segmentation," 2021, *arXiv:2102.04306*.
[33] H. Cao et al., "Swin-UNet: UNet-like pure transformer for medical image segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2022, pp. 205–218.
[34] J. Chen et al., TransUNet: Rethinking the U-Net architecture design for medical image segmentation through the lens of transformers," *Med. Image Anal.*, vol. 97, 2024, Art. no. 103280.
[35] B. A. Mork, F. Gonzalez, D. Ishchenko, D. L. Stuehm, and J. Mitra, Hybrid transformer model for transient simulation—Part I: Development and parameters*IEEE Trans. Power Del.*, vol. 22, no. 1, pp. 248–255, Jan. 2007.
[36] "Brain tumor segmentation—Kaggle.com," Accessed: Apr. 4, 2024. [Online]. Available: https://www.kaggle.com/datasets/nikhilroxtomar/brain-tumor-segmentation
[37] "Brain tumor segmentation dataset—Kaggle.com," Accessed: Apr. 4, 2024. [Online]. Available: https://www.kaggle.com/datasets/atikaakter11/brain-tumor-segmentation-dataset

MIR NAFIUL NAGIB received the B.Sc. degree in software development from Independent University, Dhaka, Bangladesh. He is currently working toward the M.Sc. degree in information technology with a major in Software Design and Management with the Washington University of Science and Technology (WUST), Alexandria, Virginia, VA, USA. His research interests include machine learning, deep learning, ML models for predictive cybersecurity threat detection in cloud systems, and deep learning approaches for efficient tumor segmentation in medical imaging.

RAHAT PERVEZ received the B.Sc. degree in computer science and engineering from North South University, Dhaka, Bangladesh. He is currently working toward the M.Sc. degree in cybersecurity with Bay Atlantic University, Washington, D.C., USA. His research interests include artificial intelligence, deep learning, explainable multilingual sentiment analysis using graph attention networks, and leveraging quantum computing for cryptocurrency market analysis.

AFSANA ALAM NOVA received the MBA degree in marketing from Independent University, Dhaka, Bangladesh. She is currently working toward the M.Sc. degree in information technology with a major in Information Technology and Management with the Washington University of Science and Technology (WUST), Alexandria, Virginia, VA, USA. Her research interests include artificial intelligence, machine learning, stock price prediction and risk assessment, and deep learning frameworks for predicting customer behavior.

HADIUR RAHMAN NABIL received the B.Sc. degree in computer science and engineering (CSE) with a major in information systems from the American International University-Bangladesh (AIUB), Dhaka, Bangladesh, in 2024. He was actively involved in competitive programming and participated in various programming competitions, notably the ACM International Collegiate Programming Contest Preliminary (ACM-ICPC) and Intra-University Programming Contests. Proficient in multiple programming languages such as C++, Python, R, Java, and PHP, he also has experience with technologies like TensorFlow, ASP.NET, MATLAB, and MySQL. He is currently a Research Assistant (RA) with the Advanced Machine Intelligence Research laboratory, where he has worked on multiple research projects published in prestigious journals and conferences. His research interests primarily revolve around artificial intelligence, deep learning, and medical imaging.

ZEYAR AUNG (Senior Member, IEEE) received the Ph.D. degree in computer science from the National University of Singapore, Singapore, in 2006. From 2006 to 2010, he was a Research Fellow with the Institute for Infocomm Research (I2R), Agency for Science, Technology, and Research (A*STAR), Singapore. In 2010, he joined the Masdar Institute, which later became a part of Khalifa University, UAE, as an Assistant Professor. He is currently an Associate Professor with the Department of Computer Science, Khalifa University, Abu Dhabi, UAE. He is also a member of the Center for Secure Cyber-Physical Systems (C2PS). His current research interests include data analytics, machine learning, and their applications in various domains, such as health care, cybersecurity, social media, financial systems, and environmental science. He is an Associate Editor of Journal of *Ambient Intelligence and Humanized Computing* (Springer Nature) and an Academic Editor of *PLOS ONE*.

M. F. MRIDHA (Senior Member, IEEE) received the Ph.D. degree in AI/ML from Jahangirnagar University, Dhaka, Bangladesh, in 2017. He is currently working as an Associate Professor with the Department of Computer Science, American International University-Bangladesh (AIUB), Dhaka, Bangladesh. He was an Associate Professor and the Chairperson with the Department of CSE, Bangladesh University of Business and Technology. He was also the CSE Department Faculty Member with the University of Asia Pacific and as a Graduate Head, from 2012 to 2019. His research experience, within both academia and industry, results in more than 120 journal and conference publications. His research was contributed to the reputed journal of Scientific Reports (Nature), Knowledge-Based Systems, Artificial Intelligence Review, IEEE ACCESS, Sensors, Cancers, and Applied Sciences. For more than ten years, he has been with the master's and undergraduate students as a supervisor of their thesis work. His research interests include artificial intelligence (AI), machine learning, deep learning, natural language processing (NLP), and Big Data analysis. He has been a Program Committee Member in several international conferences/workshops. He was an Associate Editor for several journals including *PLOS ONE* journal. He also has been a Reviewer for reputed journals and IEEE conferences like HONET, ICIEV, ICCIT, IJCCI, ICAEE, ICCAIE, ICSIPA, SCORED, ISIEA, APACE, ICOS, ISCAIE, BEIAC, ISWTA, IC3e, ISWTA, CoAST, icIVPR, ICSCT, 3ICT, and DATA21.