IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 7 June 2025, accepted 25 June 2025, date of publication 4 July 2025, date of current version 5 August 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3585926

RESEARCH ARTICLE

ABI-Net: Attention-Based Inception U-Net for Brain Tumor Segmentation From Multimodal MRI Images

EVANS KIPKOECH RUTOH[1], QIN ZHIGUANG[1], JOYCE C. BORE-NORTON[2], AND NOOR BAHADAR[3]

[1]School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China
[2]Department of Systems and Analytics, Cleveland Clinic, Cleveland, OH 44195, USA
[3]Key Laboratory of Molecular Epigenetics of the Ministry of Education (MOE), Northeast Normal University, Changchun, Jilin 130024, China

Corresponding author: Evans Kipkoech Rutoh (evansrutoh@gmail.com)

This work was supported by the National Natural Science Foundation of China through the Project Key Theories and Techniques of Privacy and Security Based on Data Life-Cycle under Project 61520106007.

ABSTRACT Magnetic Resonance Imaging (MRI) is widely used for glioma evaluation, but manual segmentation is impractical due to the large data volume. Automated techniques are necessary for precise clinical measurements. U-Net has shown promise in volumetric segmentation, but brain tumor segmentation remains challenging due to tumor diversity in type, location, and structure. This study introduces ABI-Net, an advanced U-Net variant integrating Attention-based Inception blocks for improved segmentation of brain tumor sub-regions in 3D multimodal MRI images. ABI-Net leverages the Inception module for spatial feature extraction and an attention mechanism to enhance cancerous region detection. Trained on the BraTS 2020 dataset, ABI-Net achieved dice scores of 0.8354, 0.8505, and 0.8782 for enhancing tumor (ET), tumor core (TC), and whole tumor (WT), respectively, outperforming state-of-the-art models. On the validation dataset (125 patients without segmentation masks), ABI-Net obtained average dice scores of 0.8189, 0.8401, and 0.8673 for ET, TC, and WT. ABI-Net provides an accurate, efficient solution for automated brain tumor segmentation, with significant potential for clinical applications, including diagnosis, treatment planning, and patient monitoring.

INDEX TERMS Attention mechanism, brain tumor segmentation, inception blocks, U-Net, clinical applications.

I. INTRODUCTION

The human brain is the most intricate organ of the human body, which contains billions of nerve cells that control vital functions, including intelligence, emotion regulation, physical movement, and interpretation of sensory data [1]. The smooth and proper functioning of all these tasks is directly connected to the brain's health; however, sometimes, the brain starts acting abnormally owing to different health issues. Alzheimer's disease, Parkinson's disease [2], [3], Brain Tumors, and Amyotrophic Lateral Sclerosis (ALS) are some of the most prevalent disorders of the brain [4], [5]. Among these, brain tumor is considered the most fatal brain disease, which initiates due to anomalous brain cells which could affect the neural system and the surrounding brain's healthy tissues [6]. The World Health Organization (WHO) survey depicts that 330000 people are diagnosed with this fatal disease every year. According to the severity level, origin, and speed of growth, brain tumors are categorized into several types, among them, Glioma is the most prevalent and severe type, which can be further divided into high-grade (HGG) and low-grade gliomas (LGG) [7]. The first one is the more belligerent type with a low survival rate and thus requires immediate treatment, while the second one spreads slowly and has a lifespan of several years. Early diagnosis can save significant lives by planning medical intervention and

The associate editor coordinating the review of this manuscript and approving it for publication was Claudia Raibulet.

treatment through accurate segmentation of brain tumors [8], [9], [10].

The diagnostic techniques for brain tumors primarily include biopsy, chemotherapy, radiotherapy, and medical imaging examination. Diagnosing a brain tumor is usually done manually by experts by using one imaging technique, including single-photon emission computed tomography (SPECT), computed tomography (CT), magnetic resonance imaging (MRI) and positron emission tomography (PET). CT scans can provide precise details about the structures of the tissues by using non-invasive diagnostic imaging techniques [11], but sometimes CT images ignore crucial information; therefore, experts prefer MRI for capturing a detailed and in-depth view of tissues and structures of the brain. MRI uses strong radio and magnetic waves to capture images that contain images of the T1, T1ce, T2, and FLAIR modalities. T1 modality is used to examine healthy tissues, T1ce images highlight tumor boundaries, while T2 has high-intensity images due to the presence of fluids, and FLAIR images are used to find abnormal areas [12]. Brain tumor segmentation is the process of distinguishing the healthy portion from the tumorous region; however, the task of manual segmentation becomes challenging and time-consuming for experts due to the heterogeneous nature of the MRI images [13].

The development of an effective method that can separate the tumorous region from other sub-regions and healthy tissues present in the human brain is the main difficulty in the sector of medical imaging classification and segmentation, notably for the problem of brain tumor segmentation. Therefore, relying on a single segmentation methodology is insufficient for accurately classifying the tumorous region. The gradient or weights of the local contextual information in CNN-based algorithms will vanish or practically equal zero as one moves deeper into the network. Low-level features, on the other hand, are equally crucial when predicting the segmentation mask since they provide crucial details regarding the tumorous boundaries, position, and size of the tumorous cells. The heterogeneous character of tumorous cells, which can occur in any position underneath the brain tissues with different sizes, appearances, and shapes, makes it difficult to segment sub-regions of the brain tumor [14], [15]. High computing resources are one of the primary issues with this problem since 3D deep learning problems need more resources than 2D models [16].

Furthermore, brain tumor segmentation is a highly class imbalance challenging problem, with approximately 61% of the tumor-containing area belonging to the WT class, 24% to the TC class, and 15% to the ET class [17]. The accuracy and time can be improved by automating the process of tumor detection and segmentation from MRI images. Several automatic approaches, including deep learning (DL) [18] and machine learning (ML) techniques, can be used for segmenting brain tumor [19], [20]. Traditional ML techniques rely on manual feature engineering, while DL techniques automatically extract and learn from features. Most of the SOTA using DL techniques for the automatic segmentation of brain tumor regions faced challenges of the high disparity of structure, location, and shape of tumors. This study represents a novel ABI-Net method to overcome all these challenges that utilized attention inception blocks at the contracting path of the encoder to learn diverse and comprehensive information and convolutional blocks at the expanding path to form a precise segmentation mask. The key contribution of this study is as follows:

- An Attention-based Inception U-Net (ABI-Net) architecture is designed for accurate and efficient segmenting tumor from multimodal 3D MRI brain images.
- The Inception block is used in this study to replace the original convolutional blocks of the U-Net model. A 5×5 convolutional filter's equivalent receptive field is obtained using two 3×3 filters and max pool operation in the Inception block, which considerably decreases the amount of computation required in the architecture's deeper layers.
- Extensive experiments were conducted on the proposed ABI-Net model, as well as on other variants of 3D U-Net models, to assess the performance of the proposed model. These experiments were carried out using the benchmark BraTS 2020 training dataset and evaluated on the validation dataset.
- This research makes a contribution to the domain of medical image analysis by demonstrating the potential of ABI-Net architecture for segmentation tasks.

The remainder of this research study is arranged as follows. Section II reviews related work on brain tumor segmentation applying DL methods. Section III presents the dataset used for training and evaluation with the novel ABI-Net architecture to segment tumor-containing areas of the brain by using multimodal MRI images. Section IV covers the experimental setup, performance evaluation metrics, and the outcomes of the presented method in detail, along with a performance evaluation and ablation study of the proposed model with cutting-edge segmentation models. Afterwards, in Section V, we summarize the contributions of this study and explore prospective research plans for the future.

II. RELATED WORK

The earliest accurate diagnosis and treatment of brain tumors is essential to prevent the loss of human life. Before the advancement of DL techniques, methods that were based on classical ML were being used. These ML-based methods depended on the manual handcrafted feature vectors generated to learn distinct representations of brain tumors from the MRI images. These feature vectors were then passed to the ML classification algorithms like Support Vector Classifier (SVC), K-Nearest Neighbor Classifier (KNN), etc., for classification and segmentation of tumorous cells from the healthy tissues. In the medical imaging field, there has been a rise in the use of DL architectures for accurate and reliable diagnosis. These techniques can automatically extract local

and as well as global feature maps from the input images, providing a powerful tool for medical professionals.

A. SOTA METHODS ON PATCH-WISE STRATEGY

Multimodal MRI-based brain tumor segmentation provides comprehensive information about the tumorous sub-region and assists medical specialists in the timely diagnosis and surgery planning of Gliomas. Most of the SOTA studies utilized individual modalities because of insufficient memory and computing power. Patch-wise based segmentation of brain tumor techniques was proposed to train the model on small patches with labels to correctly identify brain tissues from healthy tissues. Ballestar et al. [21] presented a patch-wise strategy to cope with the issue of high memory consumption based on the 3D V-Net model. Researchers also introduced the voxel-wised uncertainty measure to check the confidence in the predicted results. Their developed model successfully incorporated the test-time data augmentation and dropout to overcome the memory requirements and better performance using the dataset of BraTS 2020. In the work of Rehman et al. [22], the researcher presented the modified U-Net named BU-Net for brain tumor segmentation using 2D patches extracted from the 3D volumetric image. They replaced some of the convolution blocks of the U-Net architecture that contain residual and wide context blocks to enhance the local and global characteristics and to fulfill the gap between shallow and deep layers. This approach helps them to overcome the need for high-power resources. The results improved the performance of the model by 1.6% in comparison to other techniques while exploiting the BraTS 2017 and 2018 datasets. Lefkovits et al. [23] utilized and fine-tuned different pre-trained CNN architectures for multimodal segmentation of brain tumors using AWS SageMaker. The study implements the random search technique for optimizing parameters to increase the segmentation accuracy on MRI images. Finally, the study proposed the weighted average ensemble model of six pre-trained models. The BraTS 2017-2020 dataset was taken for training and evaluating of the model. However, these CNN models have less ability to learn semantic and global contextual information. Whereas transformers recently have successfully achieved fame because of the self-attention system for learning global contextual details. Zhang et al. [24] proposed SwinBTS, a swin transformer for brain tumor segmentation to preserve semantic and global contextual response by using a decoder and encoder. The model uses the convolutional operation for up sampling and down sampling and the model was evaluated using BraTS 2019, 2020, and 2021 datasets, and the suggested model achieved dice scores of 83.21, 84.75, and 91.83 for ET, TC, and WT classes on the BraTS 2021 dataset. Tripathi et al. [25] introduced a CNN model to segment brain tumor sub-regions from multimodal MRI images. The proposed model incorporates a spatial and channel attention mechanism, which allows for the preservation of the feature map across multiple modalities. This integration enhances the model's ability to accurately identify and segment the tumor sub-regions in the brain.

B. SOTA METHOD ON CLASS-IMBALANCE ISSUES

Segmentation of brain tumors is a high-level imbalance research topic, and a limited dataset is available to carry out experiment on the DL networks, many studies try to solve this problem using different solutions. Shomirov et al. [26] presented a 3D U-Net network to improve the performance of segmentation by applying the proposed weighted focal loss (WFL) function. WFL helps to solve the issue of class disparity by assigning greater weights to the minority class and lower weights to the majority class. With this technique, the proposed WFL solves the pixel degradation problem, one of the significant issues in simple loss functions. Experiments were performed on the benchmark BraTS 2019 and BraTS 2020 datasets and the above technique achieved dice scores of 0.913, 0.830, and 0.815 for WT, TC, and ET, respectively, on the BraTS 2019 dataset and 0.843, 0.871, and 0.892 on BraTS 2020 dataset for TC, ET, and WT. The study of Baid et al. [14] introduced a novel approach for intratumor brain tumor segmentation by employing the 3D U-Net model. A weighted patch extraction strategy was presented to overcome the problem of class disparity between non-tumorous and tumorous areas of the brain. 3D patches were extracted from the multimodal BraTS 2018 dataset for the training purpose of the presented model. Myronenko [27] introduced a CNN model for automatically segmenting tumors from MRI images. With some variation auto-encoder branch was added by the researcher to recreate the input image by regularizing the shared decoder to handle the problem of a small dataset for training. During the experiments, the proposed technique was evaluated using the BraTS 2018 dataset. Remarkably, the proposed technique achieved remarkable results and secured the first position in the BraTS 2018 challenge. The achieved dice scores for the different tumor sub-regions were 0.9100 for WT, 0.8233 for ET, and 0.8668 for TC. Karri et al. [28] presented a ''Scale-wise Global Contextual Axile Reverse Attention Network (SGC-ARANet)'' that improves the segmentation performance on 3D MRI images to preserve the contextual information response. The proposed model simultaneously utilized all four modalities (T1, T2, T1ce, and FLAIR) during the training phase. In this study, the global multi-level guidance unit was used to save global contextual information from the MRI image and utilized a scalar multi-facted blend module that merges the multiscale contextual data into high-level features. In the end, the axile reverse module of attention was used to capture and keep attention on multi-modalities tumorous boundaries and regions. The performance of the model was examined on the BraTS 2019 and 2020 datasets.

C. SOTA METHODS ON 3D VOLUMETRIC SEGMENTATION

Applying segmentation on medical images, especially in the context of the 3D volumetric image, global as well as local features contain essential contextual data helpful for the

segmentation task. However, due to the complex tumorous boundaries and diverse nature of the tumorous regions, it is not easy to obtain contextual information from a single stage of the model. Raza et al. [29] presented a hybrid model of U-Net and deep residual model (dResU-Net) to preserve local and global contextual response. This technique helps to save the high and low-level features to improve the segmentation ability of brain tumor sub-regions. The experiments were carried out on the BraTS 2020 dataset and achieved competitive results. The work of Naser et al. [20] introduced a multiscale cross-channel residual denoising network (MLRD-net) to preserve global local and contextual information and improve the field of reception for the network. The residual blocks used pre-activation operation to propagate the signals from the encoder part to the decoder part to avoid the model from overfitting. In the final stage, a local cross-channel denoising method was used to remove the unnecessary features without reducing the dimensions. The experiments were carried out on the BraTS 2020 dataset, and the mentioned technique obtained a dice score of 0.73, 0.91, and 0.79 for ET, WT, and TC, respectively. Segmentation of brain tumors is challenging because of the visual similarities between healthy tumors and various brain tumor modalities and sub-areas. The various techniques of pre-processing can help to solve such kinds of problems. Ilhan et al. [12] introduced tumor enhancement and localization techniques with the U-Net model to enhance the performance of the segmentation procedure. The study used a histogram-based nonparametric tumor localization technique to identify the tumor-ridden areas from an MRI image. Image enhancement was applied to these localized MRI images to increase the visual appearance and differentiate the different tumors from each other. The resultant MRI image was fed to the U-Net network for the final segmentation. The presented methodology was assessed on the 2012, 2019, and 2020 datasets of the BraTS dataset. This approach lacks semantic information about the tumorous region. Magadza et al. [31] suggest a robust network to segment and handle the diffusing nature of brain tumors and to overcome high computational complexity. The proposed model utilized depth wise convolution in the U-Net network by partially replacing some convolutional blocks with depth wise convolutional blocks. The proposed model was experimented on BraTS 2020 dataset and obtained a dice score of 0.9042, 0.7745, and 0.8286 for WT, ET, and TC, respectively.

Most existing studies on brain tumor segmentation have widely used the U-Net model and modified the U-Net model for experimental purposes. However, due to the constrained computing resources, the U-Net model has not fully explored the comprehensive information in 3D multimodal MRI images. Furthermore, due to the diversity of tumors and inter and intra-visual similarity between the different tumorous regions and normal tissues and the typical growth of brain tumors inside the human brain is still challenging. Moreover, the low-level features are equally important to high-level features to make the predicted segmentation mask because the low-level features mostly include information related to lines, edges, and boundaries of the tumor-located areas, that are being lost when the model moves deeper. Therefore, a reliable and automated technique for segmenting brain tumor sub-regions is needed in the research that can accurately and timely predict the tumorous region from multimodal MRI images so that the patient's life can be saved in its initial stages. For this purpose, this research work introduced a 3D attention-based Inception U-Net model to preserve contextual information and solve the issues found in the literature. The proposed model combines the strengths of the Inception module for feature extraction and the attention mechanism for selective feature amplification.

III. METHODOLOGY

This section discusses the details of the dataset, with pre-processing steps, and the details of the proposed ABI-Net model.

Segmentation of brain tumor is a critical function in medical imaging that have a crucial role in accurately diagnosing and treating tumors of the brain. To accurately segment brain tumors it is essential to precisely determine the tumor's location, size, and shape, which in turn can help in surgical planning, treatment monitoring, and disease prognosis. The general flow of the presented approach is given in Figure 1.

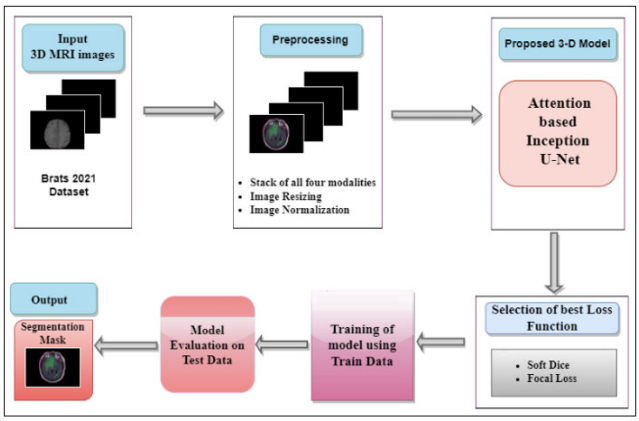


FIGURE 1. The general flow of the proposed methodology.

A. DATASET

The "Multimodal Brain Tumor Segmentation Challenge (BraTS)" is a global challenge that intends to advance the SOTA in brain tumor segmentation using multiparametric magnetic resonance imaging (mpMRI) data. This work used the publicly accessible 3D dataset of BraTS 2020 available at https://www.med.upenn.edu/cbica/brats2020/data.html. This dataset contains mpMRI scans from patients with glioblastoma and other types of brain tumors and is composed of four different MRI modalities: T1-weighted, T1-weighted with gadolinium contrast enhancement, T2-weighted, and Fluid-Attenuated Inversion Recovery (FLAIR).

The training dataset includes 76 LGG patients and 293 HGG patients. Each 3D MRI image has the dimension of

240 × 240 × 155 (where height is 240, width is 240, and 155 are the total number of slices (voxel) in each modality). The training set is subdivided into four different groups based on the tumor grade and the presence or absence of enhancing tumor regions, i.e., WT, TC, ET, and healthy. The dataset also includes manual annotations of the tumor regions by experts in the field, which serve as the ground truth for calculating the accuracy of brain tumor segmentation methods. The ground truth with the validation set is not provided, so in this research study, the training dataset is further class-wise split into validation and test sets. 80% of the total data of the training split is utilized for training of the proposed ABI-Net, 10% of the data is taken for validation purposes 10% of the data is used for the testing of the proposed model, and the dataset split was stratified. Due to the limited number of annotated samples in the dataset, we adopted 5-fold cross-validation as our primary evaluation strategy, which allows every sample to be used for testing exactly once. This was done to ensure robustness and generalizability. The full BraTS dataset was randomly split into five equal parts. In each fold, 4 partitions were used for training while 1 was held out for testing. All experiments were repeated five times, and the reported metrics (Dice, Sensitivity, Specificity) represent the average and standard deviation over the five folds. The 80:10:10 split was used primarily for training convergence analysis, while the more rigorous cross-validation ensures statistical robustness. This mitigates the limitations posed by a small standalone test set.

Table 1 provides detailed information regarding the dataset split employed in the current research.

TABLE 1. Brats 2020 training dataset.

	LGG	HGG	Total
Train Set	62	233	295
Validation Set	7	30	37
Test Set	7	30	37
Total no. of patients	76	293	369

1) PRE-PROCESSING

The BraTS 2020 challenge dataset has undergone various pre-processing steps, including all the images co-register into a single plane due to the heterogeneous nature of the different MR scanners before being shared publicly; therefore, only two pre-processing steps are applied before passing the input data to the model. Due to limited resources and to get loss-less dimensionality reduction (without decreasing the number of features), dark slices and voxels are eliminated resulting in reduced 3D dimensions from each modality, i.e., from 240 × 240 × 155 to 128 × 128 × 128 dimension in the first step. Specifically, all 3D MRI volumes were first cropped to the brain region using non-zero bounding boxes to eliminate background. Following that, the volumes were resampled to 128 × 128 × 128 using trilinear interpolation for image data and nearest-neighbor interpolation for segmentation labels to avoid label smoothing. The transformation pipeline ensured that the aspect ratio and spatial consistency were maintained while reducing memory load and standardizing input dimensions for the network. It is worth noting that it has been observed that the resizing of images has no direct correlation with the model's predicted accuracy [32]. However, the model's complexity based on the total number of parameters will increase exponentially which will make the model computationally very expensive [33]. After the image resizing, image normalization, and standardization technique is applied to all data to reduce the inhomogeneity of intensity caused by using scanners of various magnetic fields. *z-score normalization* was applied to all MRI images resulting in zero mean and standard deviation of 1, and it can be calculated by Equation (1).

$$\text{Input}_{\text{normalization}} = \frac{\text{Input} - \mu_{\text{i}}}{\sigma_{\text{i}}} \tag{1}$$

where $Input_{normalization}$ and *Input* are original and normalized input MRI images, respectively. The mean value of the input image was denoted by μ_i while σ_i represented its standard deviation. To enhance the performance of the model and capture comprehensive information from all MRI modalities, a stack consisting of T1, T2, T1ce, and FLAIR modalities was created. This stack had a shape of 128 × 128 × 128 × 4, where 128 × 128 × 128 represented the height, width, and slices of the MRI image, and 4 represented the number of MRI modalities.

Data augmentation serves as a widely adopted method to enhance the generalization capabilities of deep neural networks, essentially acting as implicit regularization. It proves particularly beneficial in scenarios where obtaining ample high-quality ground-truth data is challenging due to cost and time constraints [34].

In this study, we initially explored several augmentation techniques tailored for 3D MRI images, including affine transformations (e.g., rotation, cropping, flipping), elastic deformation, intensity transformations (e.g., brightness and contrast adjustment), and random shearing and scaling. These methods aimed to introduce anatomical and imaging variability to improve model robustness.

However, we encountered significant computational limitations during implementation. The need to process 3D images across all four modalities (T1, T1ce, T2, and FLAIR) dramatically increased memory and processing demands during training, which exceeded the capabilities of our available GPU resources. While we were able to generate augmented samples during preprocessing, the high computational load prevented us from using them in the final training pipeline.

Despite this, we acknowledge the potential of data augmentation to improve performance when sufficient resources are available. We plan to incorporate these techniques in future work, particularly in conjunction with more powerful hardware environments, to fully leverage their benefits in brain tumor segmentation tasks.

B. PROPOSED ABI-NET MODEL

The U-Net model is a widely used convolutional neural network commonly selected for image segmentation tasks. The U-Net architecture consists of a ''U'' shaped model with a decoder and an encoder, where the encoder consists of interconnected convolutional and pooling layers that down-sample the input image, and the decoder consists of a series of up-sampling and convolutional layers that restore the resolution of the image to the original size. The U-Net model uses the concatenation operation to transfer contextual information and feature maps from the contracting path to the expanding path. This process enables the classifier to make accurate predictions based on the ground truth. Although the U-Net architecture has shown success, its feature extraction method can be complex and time-consuming to execute. However, the U-Net architecture provides the flexibility to modify its encoder and decoder with modern CNN architectures, allowing for further customization and optimization.

This study introduces a novel end-to-end encoder-decoder-based modified U-Net called Attention Inception U-Net (ABI-Net), which draws inspiration from the U-Net network. The purpose of ABI-Net is to accurately segment the tumorous sub-region of the brain. The proposed ABI-Net has a five-level depth because increasing the network depth is a conventional approach to improve performance but sometimes it often leads to issues such as overfitting and increased computational complexity due to the surge in parameters. To overcome these challenges, convolutional blocks of the U-Net architecture are being replaced with Inception blocks in the first five levels of the encoder side. Leveraging the inception module, this study enables the approximation of an optimal local sparse structure by clustering the sparse matrix into denser submatrices. By substituting the original convolution modules with these inception modules, our objective is to simultaneously enhance computational performance and foster network structure sparsity. This framework facilitates the efficient processing of nonuniform sparse data, ultimately contributing to the overall efficacy of the network. Each block of the encoder part contained $3 \times 3 \times 3$ convolution and then use maxpool3D to downsampling the feature resolution. The architecture of the presented ABI-Net is shown in Figure 2.

The inception module has a unique architecture comprising parallel convolution layers with varying sizes, facilitating the effective capture of multilevel features. This inception block consists of five parallel convolutional layers and two 3D max pooling layers, one convolutional layer with a 1×1 kernel size, and the remaining four with a convolutional layer of 3×3 and 5×5 kernel sizes which are then concatenated to form an output of the inception module.

All outputs from these branches are concatenated along the channel dimension to form the final output of the inception module. It is important to note that while max pooling reduces spatial dimensions, the pooling branches are designed with appropriate stride and padding or followed by resizing operations to ensure that all feature maps have the same spatial resolution prior to concatenation. This design choice avoids dimensional mismatches and enables effective fusion of multiscale features.

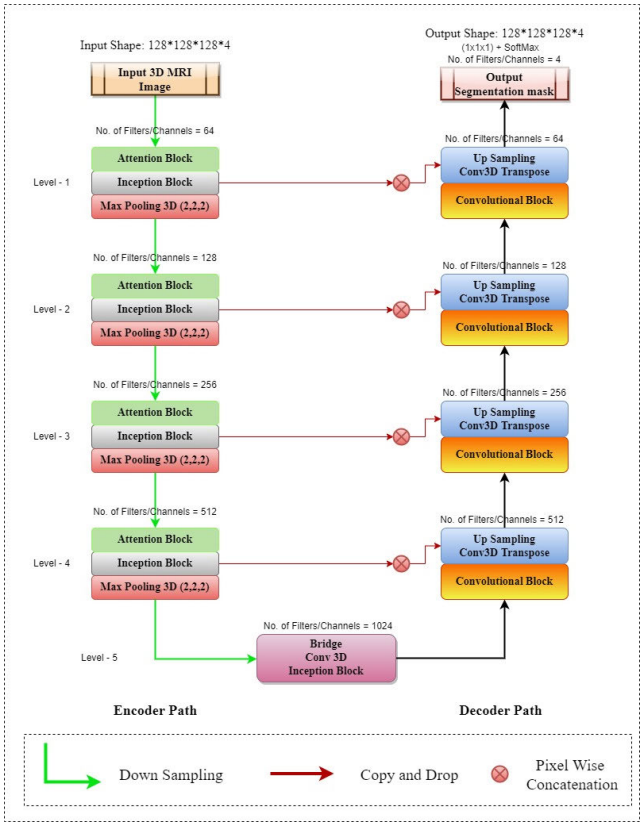


FIGURE 2. Proposed architecture of ABI-net.

By leveraging this design, the inception module achieves a balance between feature richness and computational efficiency, rendering it suitable for applications requiring multiscale feature extraction. The expanded architecture of the inception block is shown in Figure 3.

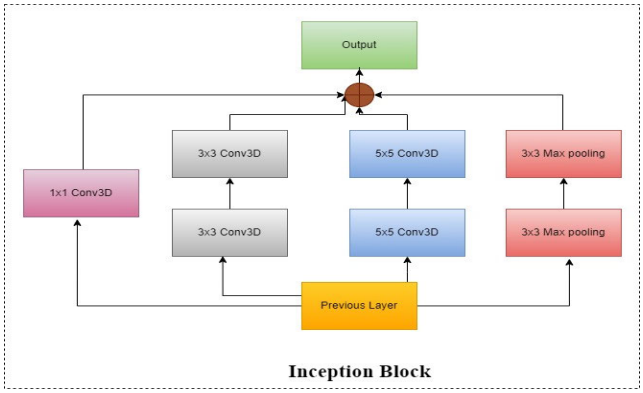


FIGURE 3. The expanded architecture of inception blocks. The circled "+" indicates concatenation of all parallel branch outputs along the channel dimension, not summation. Spatial dimensions of feature maps are preserved across branches using appropriate stride, padding, or resizing, ensuring compatibility for concatenation.

In this work, an attention block is incorporated at each level of the encoder along with the inception blocks. This integration allows for the simultaneous utilization of benefits from the attention block and the multiscale feature extraction capabilities offered by the inception block. The main motivation behind including the attention block is to enhance the overall performance of the network by effectively suppressing the activation of irrelevant and unnecessary features. Because the most common issue during feature extraction is the repetitive extraction of similar features across multiple network modules, resulting in redundant use of computational resources. To address this issue and optimize computational efficiency attention blocks are integrated that play a vital role in enhancing the accuracy of the segmentation results by capturing spatial correlations among features. By selectively recalibrating feature maps based on their significance, the attention blocks enable the network to keep attention on important information and improve discriminative capabilities. The attention mechanism used in ABI-Net is implemented as follows:

1. Input Vectors (x and g):

The attention block takes two inputs: x, the feature map from the encoder at a given level (i.e., the skip connection input), and g, the gating signal from the decoder path at the corresponding level. This gating signal carries contextual information that helps the network decide which encoder features are important at this stage of reconstruction.

2. Linear Transformations:

Both x and g are passed through separate $1 \times 1 \times 1$ convolutional layers (denoted as linear transformations) to reduce their channel dimensions to a common intermediate space. These transformations help align the semantic space of the encoder and decoder features before comparison.

3. Feature Fusion and Activation:

The transformed features are element-wise added and passed through a ReLU activation function, enabling the attention gate to learn non-linear relationships between x and g.

4. Attention Coefficient Computation:

The output from the ReLU is passed through another $1 \times 1 \times 1$ convolution layer followed by a sigmoid activation to generate an attention coefficient map which represents the relative importance of each voxel in the input feature map x.

5. Attention Weighting:

The attention coefficient map is then element-wise multiplied with the input feature map x, modulating it such that important features are emphasized and less relevant ones are suppressed.

6. Upsampling via Trilinear Interpolation:

To match spatial dimensions before concatenation with decoder features, the weighted feature map is upsampled using trilinear interpolation if needed.

7. Output Formation:

The final output of the attention block is either directly passed forward or concatenated with the decoder features to form a richer, context-aware feature representation [35].

This attention mechanism ensures that only the most contextually significant features from the encoder are passed to the decoder, enhancing segmentation accuracy—especially for small or subtle regions such as the enhancing tumor (ET) class. The expanded attention block with trilinear interpolation is shown in Figure 4.

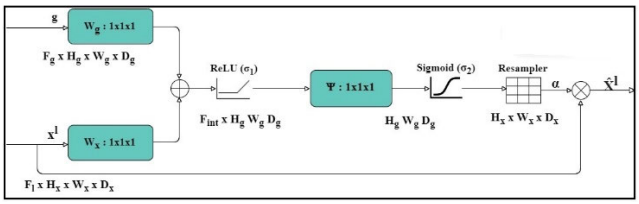


FIGURE 4. The architecture of the attention block integrated into the encoder.

The bottleneck layer of the suggested network model consists of an inception block as the bottleneck layer serves as a bridge between the decoder and encoder components of the U-Net architecture and is located at the deepest point of the encoder, where the resolution of the feature maps is the lowest. The responsibility of the bottleneck layer is to capture the most important features of the input image in a compressed representation, which the decoder can then use to produce the segmentation output mask. In the encoder part, the image as input undergoes a gradual reduction in size while experiencing an increase in the number of filters until it reaches the final block of the encoder. This last stage follows the same procedure as the preceding encoder blocks, with the processed output being forwarded to the decoder blocks via the bridge inception block.

The Decoder part of the network consists of the same levels as the decoder portion with one convolutional block and up-sampling Conv3DTranspose present in each level. Each convolutional block executes the Convolutional operation, and batch normalization process followed by the Leaky RELU activation function. The encoder blocks serve the purpose of extracting low-level semantic features from the input image in the top blocks, while the lower blocks focus on extracting high-level features. As input moves through the decoder blocks, the opposite operation occurs, where the actual size of the image is reconstructed using the up-sampling technique. The expanded architecture of convolutional blocks is present in Figure 5.

To facilitate information flow and aid in the backpropagation process for loss calculation, Skip Connections are incorporated from the Inception blocks of the encoder side to the Convolutional blocks of the decoder. These connections enable the network to retain high-level semantic image information and promote the exchange of features between the decoder and encoder blocks. Integrating skip connections within this model enhances the overall performance and information flow throughout the network. This contributes to improved feature preservation, and efficient gradient propagation, and ultimately aids in better loss computation.

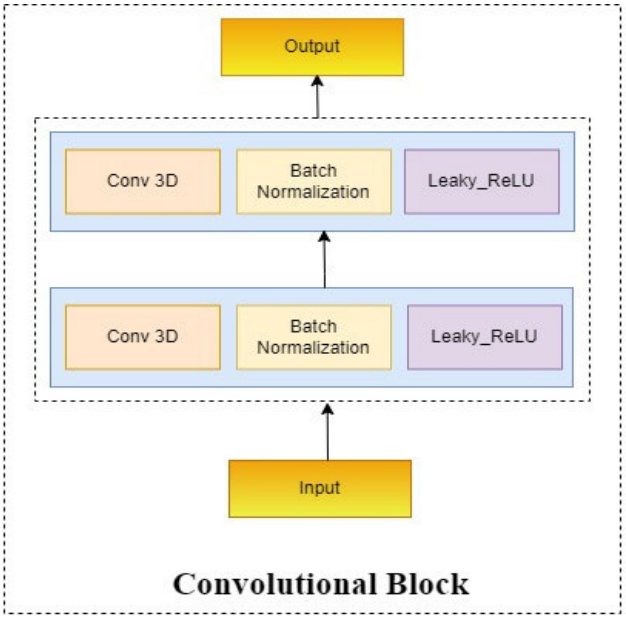


FIGURE 5. The architecture of the attention block integrated into the encoder.

C. LOSS FUNCTION

The soft dice loss function was selected and utilized for the training of the presented ABI-Net model. This loss function is derived from the Dice coefficient, which calculates the correlation between two sets. During segmenting brain tumor, it allows the delineation of regions containing tumors. There are high chances of achieving precise segmentation by employing a modified version of the dice coefficient as a loss function. This loss function can control false positives and false negatives while promoting a substantial overlap between predicted and ground truth segmentation output. The main reason behind the selection of the soft dice loss function is to minimize the soft dice loss during training, and the model aims to enhance the segmentation accuracy and capture the boundaries of the brain tumor effectively. The extensive empirical study, confirms the efficiency of this loss function in improving the performance of the U-Net-based architecture, leading to accurate segmentation of sub-regions of brain tumors. The mathematical notation for the Soft Dice Loss is given in Equation (2).

$$\textit{Soft Dice Loss}\,(x, y) = 1 - \frac{1}{N}\sum_{c=1}^{c}\frac{2X\sum_{mn} x_{cmn}y_{cmn} + \epsilon}{\left(\sum_{m,n} x_{cmn}^2 + \sum_{m,n} y_{cmn}^2\right) + \epsilon} \quad (2)$$

where x and y are the segmentation masks for the predicted and ground truth images, c is the number of classes (usually foreground and background), m and n are the pixel coordinates, and N is the total number of pixels. It is a minor constant inserted to prevent division by zero. The soft dice loss function, which is frequently used in image segmentation applications, quantifies the overlap between the predicted segmentation mask and the ground truth mask.

IV. RESULTS AND DISCUSSION

This section discusses the details related to the evaluation metrics used to examine the presented ABI-Net model's performance. Furthermore, the results obtained through the proposed technique with detailed discussion are discussed in subsequent sub-sections.

A. EVALUATION MEASURES

In the current study, the performance of brain tumor segmentation methods is evaluated using metrics such as the Dice score, specificity, and sensitivity. These metrics are employed to measure and assess the accuracy and effectiveness of the segmentation methods. The Dice score, also known as the F1 score, is a standard measure used to assess the performance of brain tumor segmentation. It calculates the overlap between the predicted segmentation mask and the ground truth mask. The Dice score ranges from 0 to 1, with a score of 1 indicating a perfect overlap between the predicted and ground truth masks. The Dice score is computed by Equation (3), as shown at the bottom of the next page, whereas the sensitivity calculates the ratio of true positive cases that are accurately identified by the segmentation method. It measures the ability of the segmentation approach to identify the tumor region in the brain correctly. Sensitivity is computed by Equation (4).

$$\text{Sensitivity} = \left(\frac{\text{True positive})}{\text{True positive} + \text{False negative}}\right) \quad (4)$$

Specificity computes the ratio of true negative cases that are accurately identified by the segmentation method. It measures the ability of the segmentation technique to identify the healthy regions in the brain correctly. Specificity is computed by Equation (5).

$$\text{Specificity} = \left(\frac{\text{True negative}}{\text{True negative} + \text{False positive}}\right) \quad (5)$$

B. IMPLEMENTATION DETAIL

In this study, the BraTS 2020 dataset was utilized to train and develop a model using the Python programming language with TensorFlow Library and Keras as the backend. The model underwent training for 100 epochs, employing an ADAM optimizer with a learning rate of 0.0005. The dataset was divided into training (80%), validation (10%), and testing (10%) sets using a class-wise split approach. All the experiments were conducted on a Google Colab Pro (GCP) server equipped with a 35 GB v100 GPU and 25 GB High RAM. Various experiments were performed on the proposed method to determine the optimal combination of hyperparameters. The learning rate and batch size were also adjusted in different experiments to identify the most suitable values. Overall, the proposed model was developed with careful consideration of hyperparameters and limitations in computational resources. The details related to the hyperparameters that were selected in the final experiments for the implementation of the presented model are mentioned in Table 2. A more comprehensive overview of the model

complexity and computational performance is provided in Section IV-F.

TABLE 2. Hyperparameter values details.

HYPERPARAMETER	VALUE
Learning Rate	0.0005
Input Shape	128x128x128x4
Batch Size	4
Output Shape	128x128x128x4
Loss Function	Dice Loss Function
No. of the Epochs	100
Optimizer	ADAM
Hidden Layer Activation Function	ReLU
Output Layer Activation Function	Softmax
Dropout	0.2

C. RESULTS OF THE PROPOSED ABI-NET MODEL

In this study, a novel ABI-Net is introduced for the efficient and timely diagnosis of sub-regions affected by brain tumor. To evaluate the performance of the model, the widely used dice score metric was chosen, which calculates the ratio of overlap between the ground truth and predicted segmentation using the benchmark BraTS 2020 dataset, which is considered a well-established dataset in the domain of medical image analysis. The suggested model attains highly promising results in segmenting different tumor areas. Specifically, on a test set the dice scores achieved by the model for the ET, WT, and TC, regions were 0.8354, 0.8782, and 0.8505, respectively. These scores indicate a significant overlap between the ground truth and predicted segmentation outputs, implying the model's proficiency in extracting the spatial extent of the tumors. The outstanding performance of the proposed model in accurately and efficiently segmenting the different tumor regions indicates great promise for clinical purposes in medical image analysis. The insights gained from this study can contribute to the advancement of the field and provide a solid foundation for future research. The quantitative results achieved from ABI-Net are given in Table 3.

D. ANALYSIS OF RESULTS AND DISCUSSION

This study proposes a novel technique that utilizes a combination of the U-Net model having attention and Inception modules for brain tumor segmentation. Our experimental outcomes indicate that incorporating Inception modules in the contracting path of the U-Net enhances the accuracy of sub-region tumor segmentation. It has been observed that the segmentation performance is improved using multiple convolutional filters with different sizes within each Inception module. These filters can capture contextual information at different scales and preserve it throughout the learning cycle, thereby improving the segmentation accuracy. While on the other hand, the attention mechanisms help the model selectively attend and focus on the essential regions of the MRI image, especially in the case of the ET class, which was the most difficult class to segment. The qualitative results attained through the suggested model are shown in Figure 6. The results indicate that the suggested model is successfully able to segment the tumorous region from the multimodal 3D MRI image.

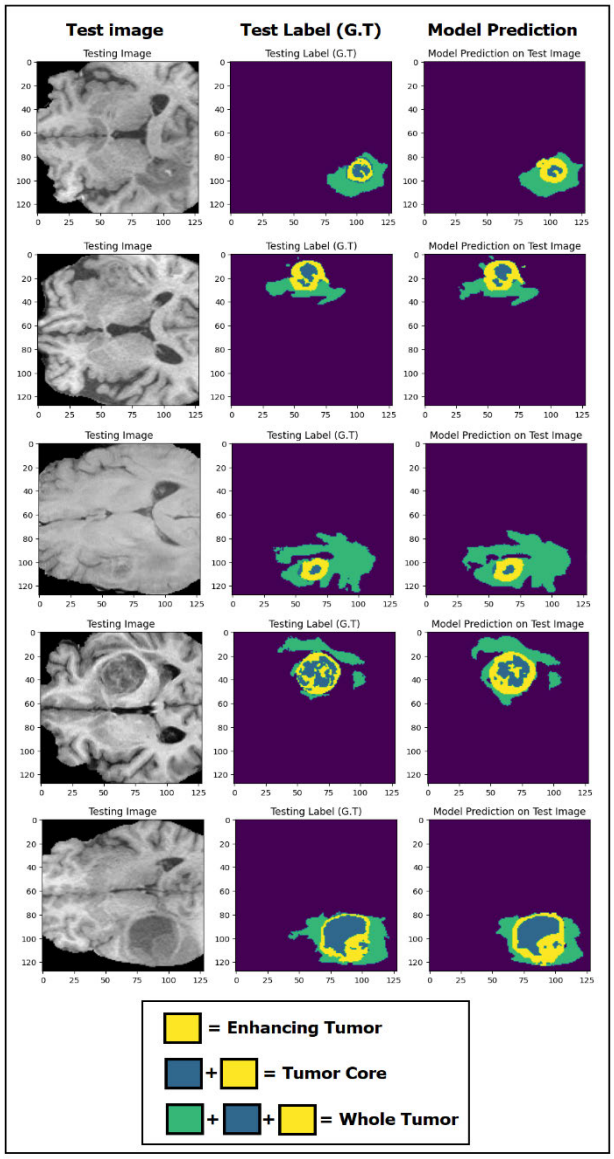


FIGURE 6. Visual outcomes of the proposed model on test MRI sample captured from an axial view.

Getting comprehensive insights into model performance for brain tumor segmentation training vs validation loss curve

$$\text{Dice score} = \left(\frac{2 * (\text{Intersection between predicted and ground truth masks})}{\text{Size of predicted mask} + \text{Size of ground truth mask}} \right) \quad (3)$$

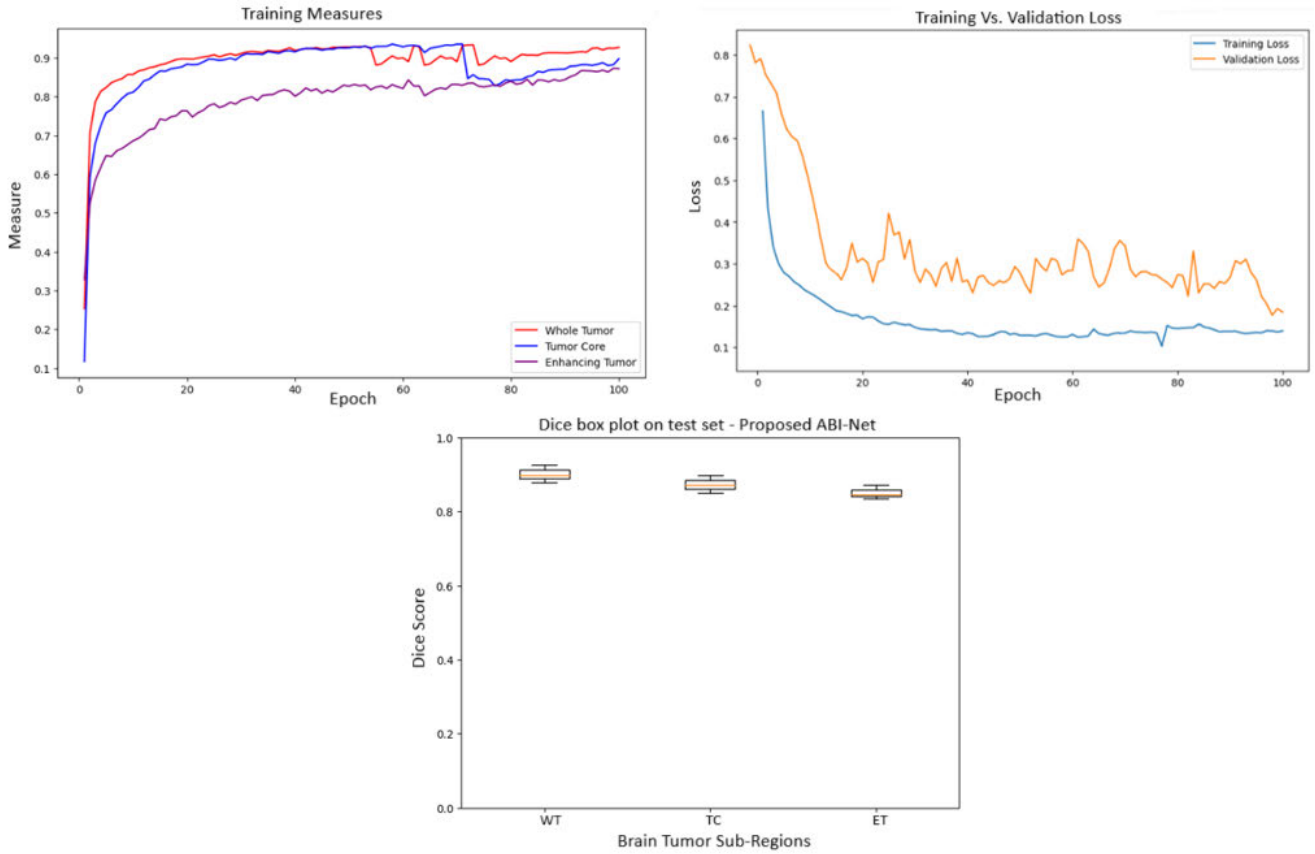


FIGURE 7. Performance metrics during model training and evaluation. (a) Dice score during training for all three tumor subregions (WT, TC, ET). (b) Training and validation loss curves over epochs. (c) Dice score boxplot on the test set for each tumor subregion.

is an important evaluation measure. This curve serves as a valuable tool in getting the model's effectiveness during the training and validation process. This analysis is instrumental in identifying whether the model is underfitting, overfitting, or achieving desirable balanced results. The results unveiled outstanding performance, due to a consistent decrease in loss throughout the training phase. Notably, the validation loss passes nearby to training loss, indicating the model's ability to identify unseen data effectively. This observation provides compelling evidence that the model successfully learned the crucial features and patterns associated with brain tumor segmentation. The convergence of the training and validation losses highlights the model's robust generalization capability and the findings from this evaluation measure confirm the model's efficacy in accurately segmenting brain tumors, offering significant promise for diagnostic and treatment planning purposes in the future. The training dice score for all three classes is shown in Figure 7(a) and the training and validation loss of the model during the model's training is shown in Figure 7(b).

It is clearly visible from Figures 7(a) and 7(b) that the model's performance is being improved as the training proceeds, and the loss is gradually reduced. The Figure 7(c) displays a boxplot representing Dice scores obtained from the evaluation of the proposed ABI-Net model on the test dataset, specifically focusing on brain tumor sub-regions. Notably, the Dice scores exhibit varying levels of performance across these sub-regions.

In the analysis of these results, it is evident that the highest Dice scores were observed within the WT region, followed by the TC region, and finally, the ET region. This discrepancy in performance reflects the inherent heterogeneity and complexity associated with the different sub-regions of brain tumors. The substantial variance observed in the Dice scores further underscores the model's capability to perform more accurate segmentations within the WT and TC regions compared to the ET region. This can be attributed to the fact that the ET region typically exhibits greater heterogeneity and noise levels, making it a more challenging area for segmentation. Moreover, the ET region is often characterized by its relatively smaller size compared to the WT and TC regions, which introduces an additional layer of complexity in achieving precise segmentation. It is noteworthy that, despite the lower Dice scores observed in the ET region, the proposed ABI-Net model demonstrates commendable performance by accurately segmenting brain tumors in all three sub-regions. This overall performance highlights the model's potential as a promising tool for brain tumor segmentation tasks,

TABLE 3. Quantitative results of the proposed ABI-Net model.

Evaluation Measure	Train Set			Validation Set			Test Set		
	ET	TC	WT	ET	TC	WT	ET	TC	WT
Dice Score	0.8719	0.8973	0.9267	0.8466	0.8727	0.8979	0.8354	0.8505	0.8782
Specificity	0.9676	0.9784	0.9791	0.9635	0.9771	0.9764	0.9658	0.9739	0.9752
Sensitivity	0.9739	0.9745	0.9810	0.9711	0.9678	0.9781	0.9615	0.9617	0.9634

offering valuable insights and potential clinical applications. To thoroughly assess the contribution of each architectural component, we constructed multiple ablation versions of ABI-Net by selectively removing key modules such as attention blocks and inception modules. These variants allow for a focused evaluation of how each design choice impacts the model's overall performance. The quantitative impact of these components on segmentation performance is detailed in Section IV-G.

1) RESULTS ON THE VALIDATION DATASET

The performance of the proposed ABI-Net model was also evaluated on the BraTS 2020 validation dataset. The BraTS 2020 validation set includes 125 patients with four modalities i.e., FLAIR, T2, T1, T1ce. The ground truth with the validation set was not publicly available and due to the unavailability of the validation server, the dice score on the validation set is calculated by performing inference on the validation data and generating segmentation masks using our trained model. The predictions were stored and obtained dice scores of 0.8673, 0.8401, and 0.8189 for WT, TC, and ET respectively. The results summarization of the proposed ABI-Net model on the BraTS 2020 validation set is given in Table 4.

TABLE 4. ABI-Net results on BraTS 2020 validation set.

Dataset	Total no. of the Patients	Results – Dice score		
		WT	TC	ET
BraTS 2020 – Validation Set	125	0.8673	0.8401	0.8189

The proposed model was evaluated on the validation set, where the ground truth was not publicly available. Figure 8 shows the segmentation results generated by the proposed model on the validation set. The results demonstrate that the proposed model can successfully segment the tumorous subregions.

To determine whether our proposed model and the baseline model represent distinct entities among the investigated algorithms with significantly varying performance levels, we conducted a paired sample t-test on the segmentation outcomes of the proposed 3D ABI-Net and the baseline 3D U-Net model. This test aimed to ascertain their statistical significance. Our alternative hypothesis posited that the mean performance of the ABI-Net surpasses that of the U-Net. The results, including the t-statistic and p-values, can be found in Table 5. The t-statistic was calculated as:

$$t = (\text{mean of differences})/(\text{standard deviation of differences}/\text{sqrt}(n)),$$

where $n = 5$ (for 5-fold cross-validation).

P-values were computed using Python's scipy.stats.ttest_rel function. Training and evaluation were performed five times (one for each fold), ensuring statistical robustness and fair comparison. Notably, the p-values for the WT, TC, and ET segments were all below the 0.05 threshold. Consequently, at the 0.05 significance level, the segmentation dice scores achieved by the ABI-Net were significantly superior to those of the U-Net. These statistical test findings support the notion that the proposed enhancement direction for the U-Net, incorporating inception and attention modules, is a viable and effective approach. This statistical analysis offers valuable insights into the ability of our proposed model to generalize well to new, previously unseen data.

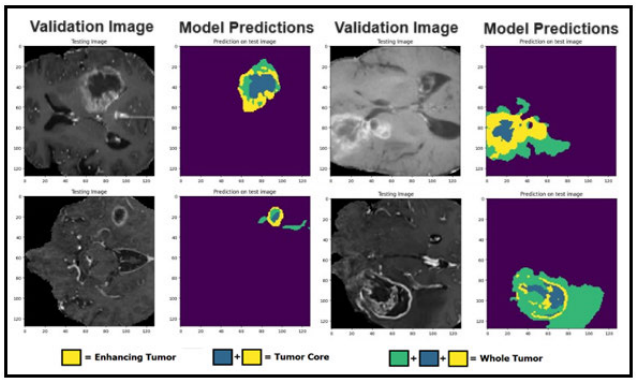


FIGURE 8. ABI-Net predictions on the validation set.

E. COMPARISON WITH STATE-OF-THE-ART METHODS

The proposed network ABI-Net achieved comparative results while using attention mechanism and inception blocks at the contracting path of the U-Net architecture. The unit

TABLE 5. Paired sample t-test on the segmentation results of proposed ABI-Net and baseline 3D U-Net.

	WT	TC	ET
t statistics	5.36817	6.08092	5.19572
p value	0.03299 (<0.05)	0.02599 (<0.05)	0.03510 (<0.05)

of attention support helps the model to concentrate on the cancerous region of the MRI image whereas the inception module helps to extract spatial features at various scales. The expanding path of the model utilized convolutional bocks to reconstruct the MRI segmentation mask. The results of the proposed model are compared with the SOTA methodologies to prove the performance and efficacy of the proposed methodology. The introduced technique successfully segments all sub-regions of the brain especially ET which was the hardest class to segment. Hence, it can be inferred that the introduced technique holds promising potential for enhancing the semantic segmentation of brain tumors. In Table 6, the comparison results of the proposed model with latest SOTA methods are given. Proposed ABI-Net model achieved superior results and efficiently segments the sub-regions of brain tumor. The latest work proposed by Raza et al. [29], introduced by the dResU-Net model by introducing the residual blocks in the contracting path of the U-Net model to overcome the issue of gradient vanishing and obtained high dice score of 86.60%, 83.57% and 80.04% for WT, TC and ET on BraTS 2020 dataset by splitting the dataset into 80:10:10 for training, validation and testing set. This study required high resources to run their experiments. By taking motivation from this study and to solve research challenges our proposed study introduced ABI-Net model to overcome the need of the high computational resources and obtained higher results than Raza et al. and achieved the dice score of 86.79%, 84.01% and 81.89% for WT, TC and ET on BraTS 2020 validation dataset, where the ground truth was not publicly available.

We have also provided a visual comparison of the segmentation outcome obtained by our experimental results with other SOTA methods. For this purpose, we compare the visual results of the SOTA methods with our proposed model. Most of the SOTA methods take test input image and produce segmentation predictions by using their proposed model and in last compare their segmentation results with ground truth and report the results based on the dice score similarity. These visual results are given in Table 7.

F. MODEL COMPLEXITY

The 3D ABI-Net model consists of 6.573 million parameters, enabling it to learn complex correlations between input MRI images and output labels. Despite its ability to capture intricate relationships, the 3D ABI-Net model is relatively simple in design. During training, our proposed model took approximately 8.83 minutes per epoch, resulting in a total training time of around 885 minutes or 14.72 hours. This training process utilized 75 compute units provided by GCP. The total computational complexity of ABI-Net is approximately 17.68 GFLOPs (Giga Floating Point Operations) per forward pass, indicating a balance between segmentation performance and computational efficiency. In terms of testing, the model demonstrated a swift duration of only 3 seconds, which is commendable for segmenting the tumorous region from cancerous 3D MRI images. For further enhancement in performance, employing more powerful GPUs and larger batch sizes can be considered.

G. ABLATION STUDY

In this study, a novel ABI-Net model is proposed to handle and automate the process of brain tumor segmentation. The proposed methodology achieved superior results as compared to the state-of-the-art methods. To prove the effectiveness of the given model an ablation study was conducted to check the impact of attention module and inception blocks on the performance of the ABI-Net model.

The attention module is incorporated at the encoder part of the model along with the inception blocks. One of the most common issues that a proposed ABI-Net model faced during the feature extraction was the repetitive extraction of similar features across multiple network modules, resulting in redundant use of computational resources. The main motivation behind the usage of the attention module at the encoder part is to improve the overall performance and efficiency of the model by effectively suppressing the activation of irrelevant and unnecessary features. The attention module successfully handles this issue resulting in high performance by capturing spatial correlations among features. The ablation study was conducted to check the impact of the attention module on the proposed model by removing these blocks from the ABI-Net model and by expanding these modules at the decoder part. It has been seen that the performance of the model was relatively poor when we trained the model without an attention mechanism. Without an attention mechanism, the proposed model obtained a dice score of 0.8551, 0.8458, and 0.8179 for WT, TC, and ET.

To further prove the effectiveness of the proposed model several experiments were performed during the ablation study. Firstly, the baseline 3D U-Net model was trained, and then we trained U-Net++ and nnU-Net to check the effectiveness of the proposed model. After the training of these

TABLE 6. Comparison of the proposed architecture with recent studies.

Study / Year	Baseline Methodology	Dataset	Dataset Type/Split	Dice Score (%)		
				WT	TC	ET
Henry et al. [36] - 2021	Two ensemble models trained separately and then output merged	BraTS-2020	Validation set	0.91	0.85	0.81
Huang et al. [37] - 2021	V-Net which includes distance Transform Decoder	BraTS-2020	Validation set	0.86	0.77	0.75
Wang et al. [38] - 2021	Novel Parallel branches-based modality-pairing end-to-end network.	BraTS-2020	Test set	0.891	0.842	0.816
Saeed et al. [38] - 2021	A network designed by taking inspiration from Residual Connections and MobileNetV2	BraTS-2020	80-20%	0.913	0.841	0.833
Parmar et al. [39] - 2021	Modified U-Net model used along with patch selection methodology	BraTS-2020	Test set	0.886	0.823	0.795
Raza et al. [29] - 2023	The U-Net model with residual network-based encoder	BraTS-2020	80-10-10%	0.866	0.836	0.800
H Liu et al. [40] - 2023	Lightweight 3D U-Net with an attention mechanism	BraTS-2020	Validation set	0.900	0.833	0.780
D Liu et al. [41] - 2023	3D self-calibrated attention U-Net	BraTS-2020	Validation set	0.905	0.821	0.781
Yousef et al. [42] - 2023	Bridged U-Net_ASPP_EVO (V2)	BraTS-2020	80-10-10%	0.907	0.816	0.78
Zhang et al. [43] - 2023	Hierarchical Multi-Scale (HMNet) model	BraTS-2020	Validation set	0.901	0.823	0.781
Akbar et al. 44]	Single level U-Net 3D with multipath residual attention model	BraTS-2020	Validation set	0.865	0.801	0.729
Proposed-Methodology	Modified U-Net model in which Attention and Inception modules are integrated into the encoder side	BraTS-2020	80-10-10%	0.878	**0.851**	**0.835**
Proposed-Methodology	Modified U-Net model in which Attention and Inception modules are integrated into the encoder side	BraTS-2020	Validation set	0.868	0.840	0.819

TABLE 7. Visual comparison of the proposed architecture segmentation results with SOTA studies.

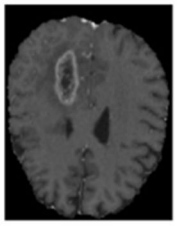
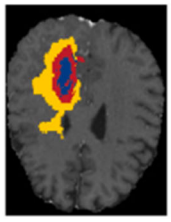
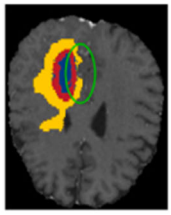
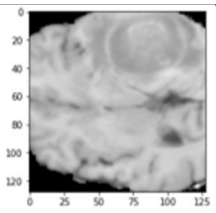
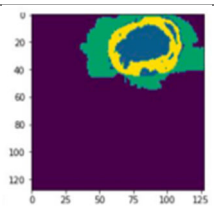
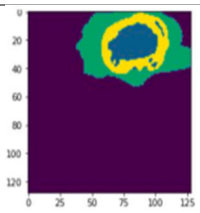
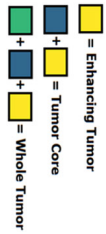
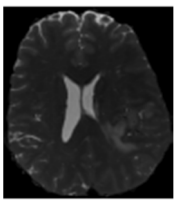
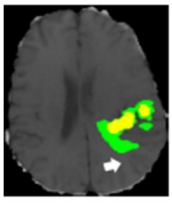
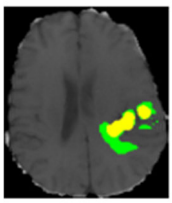
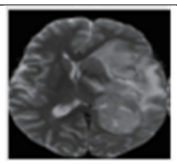
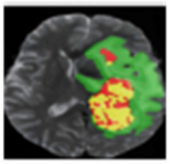
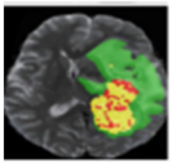
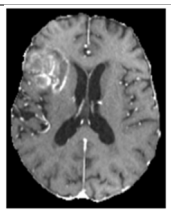
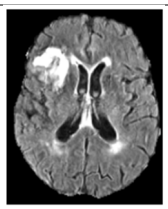
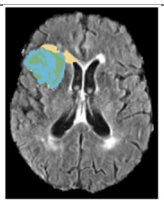
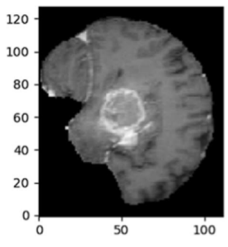
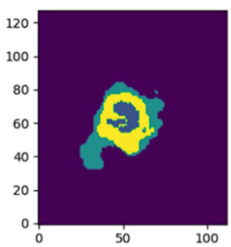
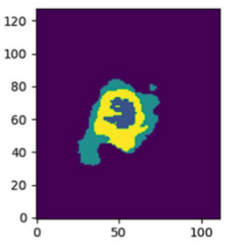
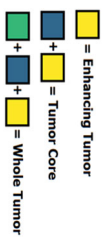
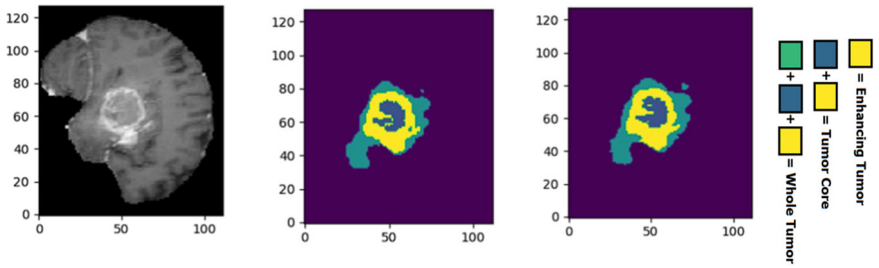
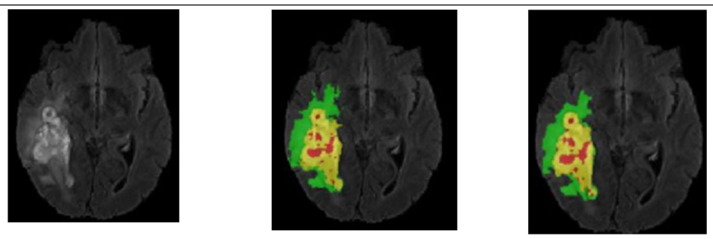
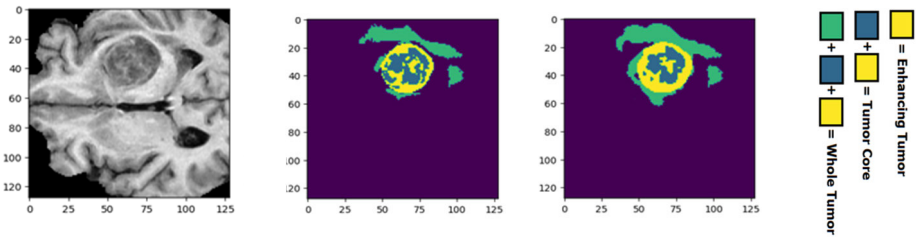
[46]	3D U-Net model with attentional feature fusion module				
[29]	3D U-Net with Residual blocks				
[47]	Hybrid of Transformer and CNN model				Green, yellow and red regions indicate ED, ET and NCR
[48]	U-Net model with symmetrical residual structure				Green represents edema; yellow indicates an enhancing tumor; and red denotes both necrosis and non-enhancing regions.
[49]	nnU-Net model with depth wise-separable convolutions				Edema is shown in yellow, necrosis in green and enhancing tumor in blue.
[50]	Multi-scale residual U-Net				

TABLE 7. *(Continued.)* **Visual comparison of the proposed architecture segmentation results with SOTA studies.**

[50]	Multi-scale residual U-Net		
[43]	HMNet model with multi-resolution feature fusion		The red area is necrosis and non-enhancing, the yellow area is enhancing tumor, and the green area is edema.
Proposed Method	**ABI-Net**		

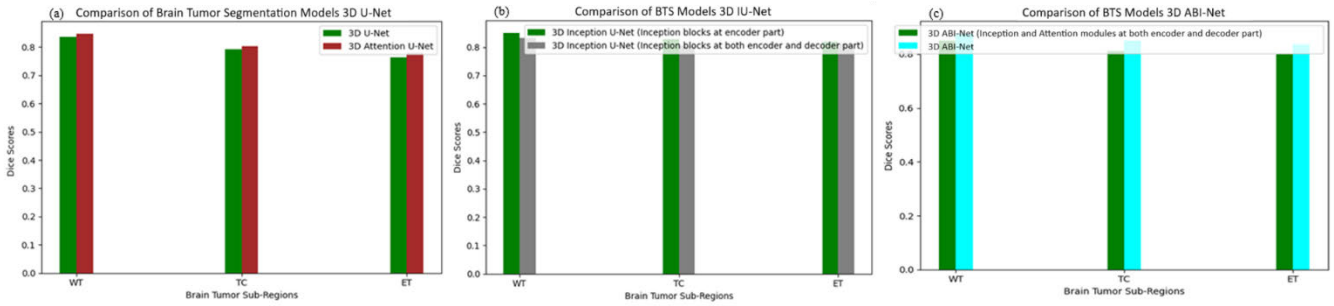


FIGURE 9. **(a) The results comparison between the 3D U-Net and the 3D Attention U-Net. (b) The results comparison between the 3D IU-Net (Inception blocks at Encoder Part) and the 3D IU-Net (Inception blocks at both Encoder and Decoder Parts). (c) The results comparison between 3D ABI-Net (Inception and Attention blocks at both Encoder and Decoder Parts) and the Proposed ABI-Net model.**

models, the attention blocks were integrated into the baseline 3D U-Net model. After that the process of integration of inception blocks at the encoder and then decoder part to check the impact and performance of the inception model into the U-Net model. In the last, inception and attention modules were expanded to both the encoder and decoder parts. The summarization of the results of the ablation study is given in Table 8.

Figure 9(a) compares the Dice scores of the baseline 3D U-Net and the proposed 3D Attention U-Net. To assess the impact of the attention module on the performance of the U-Net model, the attention mechanism was incorporated into both the encoder and decoder paths. The attention mechanism significantly improves the performance of the model by focusing on the smaller and more important regions of the tumor subregions.

Figure 9(b) shows a comparison of the Dice scores of 3D IU-Net where inception blocks were only used at the encoder part, and 3D IU-Net where inception blocks were used at both the encoder and decoder part. There is a slight variation in results because of the placement of the inception blocks. The model with inception blocks only at the contracting

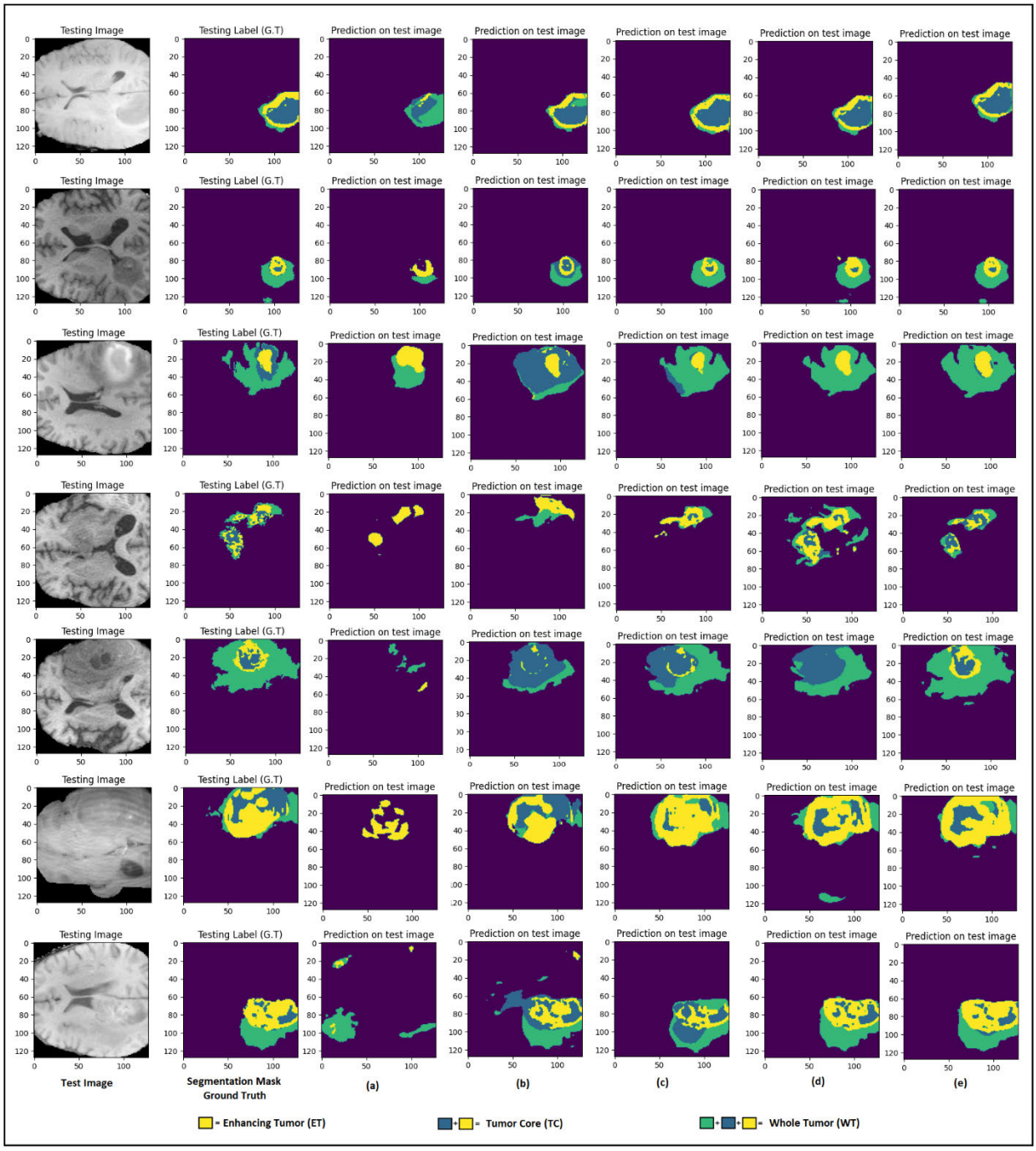


FIGURE 10. **Results comparison of segmentation models. (a) 3D U-Net predictions on Test set; (b) 3D Attention U-Net; (c) 3D ABI-Net (Without Attention Blocks); (d) 3D ABI-Net (Inception and Attention modules at both encoder and decoder part); (e) 3D ABI-Net (Proposed model).**

path obtained good results because inception blocks help the model in spatial feature extraction at multi-scale. Whereas the decoder or expanding path is only responsible for coarse feature concatenation and image reconstruction for precise segmentation.

Figure 9(c) shows a comparison of the Dice scores of two different models: 3D ABI-Net, and 3D ABI-Net (Inception and Attention modules at both encoder and decoder part). The 3D ABI-Net (Inception and Attention modules at both encoder and decoder part) model achieves the highest Dice scores on segmenting sub-regions suggesting that the proposed model is more effective at segmenting brain tumors than the other models. The main reason for the superior performance of the proposed model is the use of Inception

TABLE 8. Results of the ablation study on the test set.

Models	Results		
	WT	TC	ET
3D U-Net model - Baseline	0.8152	0.7915	0.7636
3D U-Net++	0.8215	0.8131	0.7840
3D nnU-Net	0.8488	0.8245	0.7880
3D Attention U-Net	0.8479	0.8038	0.7794
3D Inception U-Net (Inception blocks at encoder part)	0.8510	0.8276	0.8191
3D Inception U-Net (Inception blocks at both encoder and decoder part)	0.8329	0.8059	0.8117
3D ABI-Net (Inception and Attention modules at both encoder and decoder part)	0.8500	0.8117	0.8006
3D ABI-Net (Without Attention Blocks)	0.8551	0.8458	0.8179
3D ABI-Net – Proposed	**0.8782**	**0.8505**	**0.8354**

and Attention modules that enhance the efficiency and accuracy of the proposed model by learning spatial relationships between different parts of the brain tumor. Attention modules help the model to focus on the most important parts of the input image.

The proposed model exhibits notable performance, achieving Dice scores of 0.8782, 0.8505, and 0.8354 for the Whole Tumor (WT), Tumor Core (TC), and Enhancing Tumor (ET) regions, respectively. This quantitative assessment is substantiated by a visual analysis of the model's segmentation results, confirming its effectiveness in delineating brain tumor regions. The proposed ABI-Net model emerges as a strong competitor when compared to other existing models in the field. To provide a visual representation of the segmentation outcomes, we refer to Figure 10, where the results are depicted.

It is important to acknowledge that, while some minor segmentation imperfections are observed in our proposed method, these instances are markedly closer to the ground-truth labels, signifying the feasibility and efficacy of our proposed enhancement strategy. Furthermore, there exist instances in which the baseline 3D U-Net either fails or achieves only moderate success in segmenting brain tumors. In contrast, our proposed model demonstrates the capability to successfully segment such challenging cases. Notably, the incorporation of the attention module into the proposed model has led to a significant enhancement in the network's average Dice score, surpassing that of the baseline network and other models used for comparison. This observed improvement underscores the value of the attention module in refining feature extraction, with a particular focus on smaller, yet crucial, regions of interest within the brain tumor. The collective evidence supports the conclusion that our proposed ABI-Net model offers a substantial enhancement in brain tumor segmentation performance, making it a promising and competitive solution in the field of medical image analysis.

V. CONCLUSION AND FUTURE DIRECTIONS

Brain tumor segmentation is crucial for efficient prediction and planning for the treatment of brain tumors. In this research study, an ABI-Net model is introduced to segment tumorous sub-regions of the brain i.e., WT, ET, and TC. The proposed model successfully integrates the attention mechanisms and inception modules into the U-Net architecture. The attention module facilitates the model to concentrate on significant areas of the input images and handle the challenges associated with brain tumor segmentation, such as class imbalance, small object detection, and noisy images. At the same time, the inception module allows the model to extract features at various levels and be able to accurately segment tumors of different sizes and shapes, making it a reliable tool for assisting radiologists in clinical practice. The combination of these two techniques leads to a significant improvement in the segmentation accuracy of brain tumors. Brain tumor segmentation using the BraTS dataset inherently suffers from severe class imbalance, especially in the Enhancing Tumor (ET) region. To address this issue and ensure balanced learning across all tumor subregions, the proposed ABI-Net incorporates the following strategies:

Use of Soft Dice Loss:

ABI-Net is trained using soft Dice loss, which is particularly effective for imbalanced segmentation tasks. Unlike standard loss functions such as cross-entropy, soft Dice loss directly maximizes the overlap between the predicted and ground truth regions, regardless of their size. This ensures that even small regions like ET contribute significantly to the loss and gradient during backpropagation, encouraging the model to focus on these underrepresented classes.

Attention Modules for Spatial Focus:

The integration of attention blocks within the encoder enhances the model's capacity to selectively attend to spatially relevant regions. These blocks recalibrate the importance of feature maps, helping the network emphasize difficult-to-segment structures such as ET while suppressing background or irrelevant features.

Skip Connections for Feature Preservation:

Skip connections between encoder and decoder layers facilitate the transfer of spatial detail and semantic richness across the network. This continuity helps preserve small but critical features during upsampling, which is essential for accurately reconstructing underrepresented tumor regions.

Together, these architectural and training choices enable ABI-Net to effectively handle the class imbalance problem, resulting in improved segmentation performance across all tumor subregions—including the challenging ET class—as demonstrated in both the quantitative evaluation and qualitative visualizations.

The proposed approach obtained a dice score of 0.8354, 0.8505, and 0.8782 for ET, TC, and WT, respectively, using the publicly available benchmark BraTS 2020 dataset. The results indicate that the suggested approach performs better in segmenting individual regions of tumor such as WT, TC,

and ET, closer to the actual ground truth. The findings of this study reveal the potential of ABI-Net architecture for medical image analysis and segmentation tasks. In the future, the proposed architecture's performance can be enhanced by prioritizing the improvement of its interpretability and investigating its suitability for various other medical imaging tasks. Furthermore, in the future, the post-processing technique and larger datasets can be used to boost the accuracy of the system and can be deployed in real time to facilitate the medical professionals.

REFERENCES

[1] M. Bear, B. Connors, and M. A. Paradiso, *Neuroscience: Exploring the Brain, Enhanced Edition: Exploring the Brain*. Boston, MA, USA: Jones & Bartlett, 2020.

[2] J. C. Bore, B. A. Campbell, H. Cho, F. Pucci, R. Gopalakrishnan, A. G. Machado, and K. B. Baker, "Long-lasting effects of subthalamic nucleus coordinated reset deep brain stimulation in the non-human primate model of parkinsonism: A case report," *Brain Stimulation*, vol. 15, no. 3, pp. 598–600, May 2022.

[3] J. C. Bore, B. A. Campbell, H. Cho, R. Gopalakrishnan, A. G. Machado, and K. B. Baker, "Prediction of mild parkinsonism revealed by neural oscillatory changes and machine learning," *J. Neurophysiol.*, vol. 124, no. 6, pp. 1698–1705, Dec. 2020, doi: 10.1152/jn.00534.2020.

[4] B. N. Dugger and D. W. Dickson, "Pathology of neurodegenerative diseases," *Cold Spring Harbor Perspectives Biol.*, vol. 9, no. 7, 2017, Art. no. a028035.

[5] G. G. Kovacs, "Molecular pathology of neurodegenerative diseases: Principles and practice," *J. Clin. Pathol.*, vol. 72, no. 11, pp. 725–735, Nov. 2019.

[6] A. L. Vescovi, R. Galli, and B. A. Reynolds, "Brain tumour stem cells," *Nature Rev. Cancer*, vol. 6, no. 6, pp. 425–436, Jun. 2006.

[7] H.-h. Cho and H. Park, "Classification of low-grade and high-grade glioma using multi-modal image radiomics features," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2017, pp. 3081–3084.

[8] E. Bourkoula et al., "Glioma-associated stem cells: A novel class of tumor-supporting cells able to predict prognosis of human low-grade gliomas," *STEM CELLS*, vol. 32, no. 5, pp. 1239–1253, May 2014.

[9] F. Hu, X. Shi, H. Wang, N. Nan, K. Wang, S. Wei, Z. Li, S. Jiang, H. Hu, and S. Zhao, "Is health contagious? Based on empirical evidence from China family panel studies' data," *Frontiers Public Health*, vol. 9, Jul. 2021, Art. no. 691746.

[10] E. K. Rutoh, Q. Zhi Guang, N. Bahadar, R. Raza, and M. S. Hanif, "GAIR-U-net: 3D guided attention inception residual u-net for brain tumor segmentation using multimodal MRI images," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 36, no. 6, Jul. 2024, Art. no. 102086.

[11] S. Lu, B. Yang, Y. Xiao, S. Liu, M. Liu, L. Yin, and W. Zheng, "Iterative reconstruction of low-dose CT based on differential sparse," *Biomed. Signal Process. Control*, vol. 79, Jan. 2023, Art. no. 104204.

[12] A. Ilhan, B. Sekeroglu, and R. Abiyev, "Brain tumor segmentation in MRI images using nonparametric localization and enhancement methods with U-net," *Int. J. Comput. Assist. Radiol. Surgery*, vol. 17, no. 3, pp. 589–600, Mar. 2022.

[13] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in MRI images," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1240–1251, May 2016.

[14] U. Baid, S. Talbar, S. Rane, S. Gupta, M. H. Thakur, A. Moiyadi, N. Sable, M. Akolkar, and A. Mahajan, "A novel approach for fully automatic intra-tumor segmentation with 3D U-Net architecture for gliomas," *Frontiers Comput. Neurosci.*, vol. 14, p. 10, Feb. 2020.

[15] P. Pedapati and R. V. Tannedi, "Brain tumour detection using hog by SVM," Tech. Rep., 2018. [Online]. Available: https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1184069&dswid=8126

[16] M. Liu, X. Zhang, B. Yang, Z. Yin, S. Liu, L. Yin, and W. Zheng, "Three-dimensional modeling of heart soft tissue motion," *Appl. Sci.*, vol. 13, no. 4, p. 2493, Feb. 2023.

[17] I. Aboussaleh, J. Riffi, A. M. Mahraz, and H. Tairi, "Brain tumor segmentation based on deep Learning's feature representation," *J. Imag.*, vol. 7, no. 12, p. 269, Dec. 2021.

[18] J. C. Bore, P. Li, L. Jiang, W. M. A. Ayedh, C. Chen, D. J. Harmah, D. Yao, Z. Cao, and P. Xu, "A long short-term memory network for sparse spatiotemporal EEG source imaging," *IEEE Trans. Med. Imag.*, vol. 40, no. 12, pp. 3787–3800, Dec. 2021, doi: 10.1109/TMI.2021.3097758.

[19] S. Lu, S. Liu, P. Hou, B. Yang, M. Liu, L. Yin, and W. Zheng, "Soft tissue feature tracking based on deep matching network," *Comput. Model. Eng. Sci.*, vol. 136, no. 1, pp. 363–379, 2023.

[20] M. A. Naser and M. J. Deen, "Brain tumor segmentation and grading of lower-grade glioma using deep learning in MRI images," *Comput. Biol. Med.*, vol. 121, Jun. 2020, Art. no. 103758.

[21] L. Mora Ballestar and V. Vilaplana, "Brain tumor segmentation using 3D-CNNs with uncertainty estimation," 2020, *arXiv:2009.12188*.

[22] M. U. Rehman, S. Cho, J. H. Kim, and K. T. Chong, "BU-net: Brain tumor segmentation using modified U-net architecture," *Electronics*, vol. 9, no. 12, p. 2203, Dec. 2020.

[23] S. Lefkovits, L. Lefkovits, and L. Szilágyi, "HGG and LGG brain tumor segmentation in multi-modal MRI using pretrained convolutional neural networks of Amazon sagemaker," *Appl. Sci.*, vol. 12, no. 7, p. 3620, Apr. 2022.

[24] Y. Jiang, Y. Zhang, X. Lin, J. Dong, T. Cheng, and J. Liang, "SwinBTS: A method for 3D multimodal brain tumor segmentation using Swin transformer," *Brain Sci.*, vol. 12, no. 6, p. 797, Jun. 2022.

[25] P. C. Tripathi and S. Bag, "An attention-guided CNN framework for segmentation and grading of glioma using 3D MRI scans," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 20, no. 3, pp. 1890–1904, May 2022.

[26] A. Shomirov, J. Zhang, and M. M. Billah, "Brain tumor segmentation of HGG and LGG MRI images using WFL-based 3D U-Net," *J. Biomed. Sci. Eng.*, vol. 15, no. 10, pp. 241–260, 2022.

[27] A. Myronenko, "3D MRI brain tumor segmentation using autoencoder regularization," in *Proc. Int. MICCAI Brainlesion Workshop*, 2019, pp. 311–320.

[28] M. Karri, C. S. R. Annvarapu, and U. R. Acharya, "SGC-ARANet: Scale-wise global contextual axile reverse attention network for automatic brain tumor segmentation," *Appl. Intell.*, vol. 53, no. 12, pp. 15407–15423, Jun. 2023.

[29] R. Raza, U. I. Bajwa, Y. Mehmood, M. W. Anwar, and M. H. Jamal, "DResU-net: 3D deep residual U-net based brain tumor segmentation from multimodal MRI," *Biomed. Signal Process. Control*, vol. 79, Jan. 2023, Art. no. 103861.

[30] X. Chen, Y. Peng, Y. Guo, J. Sun, D. Li, and J. Cui, "MLRD-net: 3D multiscale local cross-channel residual denoising network for MRI-based brain tumor segmentation," *Med. Biol. Eng. Comput.*, vol. 60, no. 12, pp. 3377–3395, Dec. 2022.

[31] T. Magadza and S. Viriri, "Brain tumor segmentation using partial depthwise separable convolutions," *IEEE Access*, vol. 10, pp. 124206–124216, 2022.

[32] Q. Wang and Y. Yuan, "Learning to resize image," *Neurocomputing*, vol. 131, pp. 357–367, May 2014.

[33] O. Rukundo, "Effects of image size on deep learning," *Electronics*, vol. 12, no. 4, p. 985, Feb. 2023.

[34] J. Nalepa, M. Marcinkiewicz, and M. Kawulok, "Data augmentation for brain-tumor segmentation: A review," *Frontiers Comput. Neurosci.*, vol. 13, p. 83, Dec. 2019.

[35] J. Nodirov, A. B. Abdusalomov, and T. K. Whangbo, "Attention 3D U-Net with multiple skip connections for segmentation of brain tumor images," *Sensors*, vol. 22, no. 17, p. 6501, Aug. 2022.

[36] T. Henry, A. Carré, M. Lerousseau, T. Estienne, C. Robert, N. Paragios, and É. Deutsch, "Brain tumor segmentation with self-ensembled, deeply-supervised 3D U-net neural networks: A BraTS 2020 challenge solution," in *Proc. Int. MICCAI Brainlesion Workshop*, Oct. 2020, pp. 327–339.

[37] H. Huang, G. Yang, W. Zhang, X. Xu, W. Yang, W. Jiang, and X. Lai, "A deep multi-task learning framework for brain tumor segmentation," *Frontiers Oncol.*, vol. 11, Jun. 2021, Art. no. 690244.

[38] Y. Wang, Y. Zhang, F. Hou, Y. Liu, J. Tian, C. Zhong, Y. Zhang, and Z. He, "Modality-pairing learning for brain tumor segmentation," in *Proc. Int. MICCAI Brainlesion Workshop*, Oct. 2021, pp. 230–240.

[39] B. Parmar and M. Parikh, "Brain tumor segmentation and survival prediction using patch based modified 3D U-Net," in *Proc. Int. MICCAI Brainlesion Workshop*, Oct. 2021, pp. 398–409.

[40] H. Liu, G. Huo, Q. Li, X. Guan, and M.-L. Tseng, "Multiscale lightweight 3D segmentation algorithm with attention mechanism: Brain tumor image segmentation," *Expert Syst. Appl.*, vol. 214, Mar. 2023, Art. no. 119166.

[41] D. Liu, N. Sheng, Y. Han, Y. Hou, B. Liu, J. Zhang, and Q. Zhang, "SCAU-net: 3D self-calibrated attention U-Net for brain tumor segmentation," *Neural Comput. Appl.*, vol. 2023, pp. 1–13, Jan. 2023.
[42] R. Yousef, S. Khan, G. Gupta, B. M. Albahlal, S. A. Alajlan, and A. Ali, "Bridged-U-Net-ASPP-EVO and deep learning optimization for brain tumor segmentation," *Diagnostics*, vol. 13, no. 16, p. 2633, Aug. 2023.
[43] R. Zhang, S. Jia, M. J. Adamu, W. Nie, Q. Li, and T. Wu, "HMNet: Hierarchical multi-scale brain tumor segmentation network," *J. Clin. Med.*, vol. 12, no. 2, p. 538, Jan. 2023.
[44] A. S. Akbar, C. Fatichah, and N. Suciati, "Single level UNet3D with multipath residual attention block for brain tumor segmentation," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 34, no. 6, pp. 3247–3258, Jun. 2022.
[45] Z. Zhu, X. He, G. Qi, Y. Li, B. Cong, and Y. Liu, "Brain tumor segmentation based on the fusion of deep semantics and edge information in multimodal MRI," *Inf. Fusion*, vol. 91, pp. 376–387, Mar. 2023.
[46] T. Zhou, "Modality-level cross-connection and attentional feature fusion based deep neural network for multi-modal brain tumor segmentation," *Biomed. Signal Process. Control*, vol. 81, Mar. 2023, Art. no. 104524.
[47] J. Lin, J. Lin, C. Lu, H. Chen, H. Lin, B. Zhao, Z. Shi, B. Qiu, X. Pan, Z. Xu, B. Huang, C. Liang, G. Han, Z. Liu, and C. Han, "CKD-TransBTS: Clinical knowledge-driven hybrid transformer with modality-correlated cross-attention for brain tumor segmentation," *IEEE Trans. Med. Imag.*, vol. 42, no. 8, pp. 2451–2461, Aug. 2023.
[48] Z. Rahman, R. Zhang, and J. A. Bhutto, "A symmetrical approach to brain tumor segmentation in MRI using deep learning and threefold attention mechanism," *Symmetry*, vol. 15, no. 10, p. 1912, Oct. 2023.
[49] T. Magadza and S. Viriri, "Efficient nnU-net for brain tumor segmentation," *IEEE Access*, vol. 11, pp. 126386–126397, 2023.
[50] P. Li, Z. Li, Z. Wang, C. Li, and M. Wang, "MResU-net: Multi-scale residual U-net-based brain tumor segmentation from multimodal MRI," *Med. Biol. Eng. Comput.*, vol. 62, no. 3, pp. 1–11, Mar. 2024.

EVANS KIPKOECH RUTOH received the bachelor's degree in BBIT and the M.S. degree in software engineering from the University of Electronic Science and Technology of China, where he is currently pursuing the Ph.D. degree in software engineering. His research interests include deep learning, brain tumor segmentation, convolutional neural networks, attention mechanism, convolutional layers, and software development.

QIN ZHIGUANG received the Ph.D. degree from the University of Electronic Science and Technology of China (UESTC), in 1996. He is currently the Retired Dean of the School of Information and Software Engineering, UESTC, where he is also the Director of the Key Laboratory of New Computer Application Technology and the UESTC-IBM Technology Centre. His research interests include computer networking, convolutional neural networks, attention mechanism, convolutional layers, deep learning, information security, cryptography, information management, intelligent traffic, electronic commerce, distribution, and middleware.

JOYCE C. BORE-NORTON received the M.S. and Ph.D. degrees in biomedical engineering from the School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, Sichuan, China, in 2016 and 2019, respectively. She completed her postdoctoral training in neurosciences at Cleveland Clinic, Cleveland, OH, USA. She is currently a Systems Analyst with the Cleveland Clinic. Previously, she taught neuroscience at Baldwin Wallace University, Berea, OH, USA. Her research interests include brain image analysis, data science, machine learning, and artificial intelligence.

NOOR BAHADAR is advances biomedical research by merging gene editing and AI. His work spans CRISPR-based genome editing in mice, elucidating gene function, and developing deep learning models for medical image analysis. He is also involved in applying computational methods for cancer drug discovery. With interdisciplinary expertise in molecular biology and AI. He is committed to pioneering innovative, translational solutions that improve diagnostics, therapeutics, and our understanding of human disease.