Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 22 No. 3 (2026) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v22i03.58359

PAPER

Hybrid Attention and Learnable Thresholding for 3D Brain Tumor Segmentation

Salma Adel Ghali (✉), Ali El-Zaart, Lama Affara

Beirut Arab University, Beirut, Lebanon

s.ghali@bau.edu.lb

ABSTRACT

Rapid and accurate segmentation of tumors in medical images is of primary importance for cancer diagnosis and treatment designs. Though advancements have been made with auto-encoding 3D images and U-Net-like architectures, the segmentation boundary precision problem is still persistent with complex, oddly shaped tumors under constrained input conditions. Advancing boundary precision remains one of the field's most important open challenges. This study proposes an improvement to the 3D autoencoder by incorporating a hybrid model. We improve feature extraction through the addition of a hybrid attention (HA) mechanism, which combines channel-wise and learned query structural attention. We also include a differentiable learnable thresholding (DLT) layer, which provides data-driven segmentation boundary control and segmentation boundary refinement for segmentation. The model was evaluated on the Brain Tumor Segmentation (BraTS) 2020 dataset and achieved a 92.0 dice score on the whole tumor. The defined metrics serve as robust evidence for accuracy and demonstrate the improvement of the proposed model over existing state-of-the-art benchmarks. Increased precision and accuracy offered by the proposed system surpasses the current computer-assisted diagnostic systems for medical imaging. This study pushes the boundary of existing knowledge as it helps to resolve the challenging issue of pinpointing the location of tumors in 3D medical volumes.

KEYWORDS

3D medical image, tumor segmentation, deep learning, 3D autoencoder, convolutional neural network (CNN), hybrid attention (HA) mechanism, learnable thresholding

1 INTRODUCTION

The recent improvements in magnetic resonance imaging (MRI) and computed tomography (CT) scans have opened possibilities in diagnosing and treating various disorders. One disorder that poses the most challenges and worsens a person's chances of surviving and improving the quality of life during and after treatment is that of a brain tumor. One particularly difficult type of tumor is edema, which poses

Ghali, S. A., El-Zaart, A., Affara, L. (2026). Hybrid Attention and Learnable Thresholding for 3D Brain Tumor Segmentation. *International Journal of Online and Biomedical Engineering (iJOE)*, 22(3), pp. 73–93. https://doi.org/10.3991/ijoe.v22i03.58359

Article submitted 2025-08-21. Revision uploaded 2025-11-17. Final acceptance 2025-12-12.

a significant challenge. The tumors are stealthy and highly aggressive, which is why they require urgent and, in most cases, surgical intervention due to soft-tissue infiltration patterns and proximity to vital neuroanatomical structures.

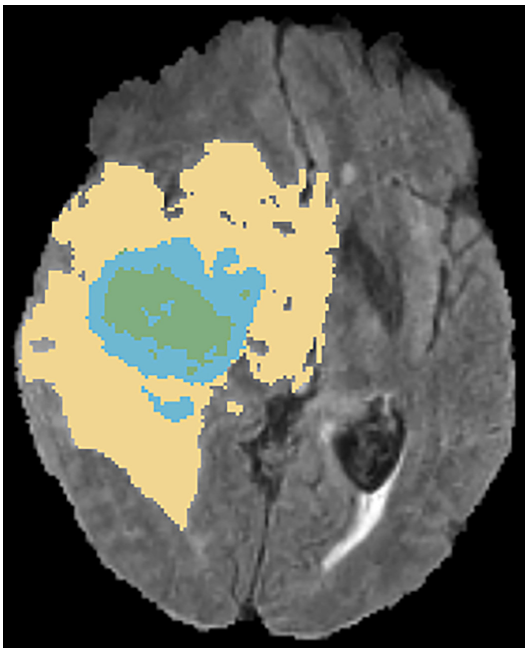

Fig. 1. Example of brain tumor regions

Consequently, accurate brain tumor segmentation, which typically involves delineating distinct sub-regions, as shown in Figure 1, is critical for several key tasks. In such cases, precise surgical planning needs to include the procedures to be performed to define the tumors in the region of interest, calculate their volumes, and define their growth patterns. It also defines surgical targets for the manipulation of the brain tumor tissue in order to enhance the postoperative prognosis, control, and follow up.

Training and refining techniques to segment these heterogeneous pairs of structures results in excessive tedium and inordinate amounts of labor, along with excessive variability across different examiners. Therefore, for the purpose of automation in the case of diagnosing and treating a patient, as well as for tracking the progression of a disease, timely and precise segmentation of brain tumors is essential. Therefore, automation of segmentation provides considerable ease and effectiveness for the planning and monitoring of disease management, particularly in the segmentation of brain tumors and the surrounding complex anatomy of its varying structures.

Medical image segmentation was historically undertaken using classic techniques like thresholding, edge detection, region-based methods, and clustering techniques. Although these approaches set an important foundation, their limitations in effectiveness largely resulted from the intricacies and the variabilities of clinical data. On the other hand, the advancement of deep learning, especially its use in Convolutional Neural Networks (CNNs) [4], [15], has become extraordinarily successful and one of the leading techniques in medical image evaluation. The 3D autoencoders, especially [14] and its iterative enhancements [2], [7], [17], have become proficient in volumetric medical image reconstruction, denoising, and segmentation. Nevertheless, the pursuit of attaining higher segmentation accuracy on irregularly shaped tumors or lesions defined by ambiguous borders in convoluted anatomical regions still poses

a challenge. Recently, there has been an increasing focus on enhancing segmentation model robustness and precision through innovative network designs, new regularization methods, and improved multi-head attention mechanisms. The inability of standard convolutional layers to overcome fixed, static receptive fields continues to challenge the adaptive weighing of prominent structures and the many different tumor morphologies in contrast and manifestation seen in brain magnetic resonance imaging.

To address these limitations, recent advancements have explored the integration of attention mechanisms [18], [19], [22] within deep learning models [16], [20], [21]. Integrating attention mechanisms allows deep learning models to focus on salient attributes of the input features, whereas traditional deep learning models 'see' elements of an input feature using fixed convolutional receptive fields. While the attention mechanisms described do improve sophisticated feature learning, the boundaries of the segmentation maps produced still require additional work. In particular, classic post-processing techniques, such as Otsu's, rely on static thresholds [23]. In contrast, our work introduces an endpoint differentiable, sigmoid-based learnable thresholding mechanism [24]. This allows the model to dynamically learn a threshold bias, attenuating, modulating, and binarizing the attention-weighted features in an efficient way that is trainable end-to-end.

This study develops a refined three-dimensional autoencoder architecture based on the standard 3D U-Net configuration, focusing on the advanced segmentation of brain tumors. Based on hybrid attention (HA) mechanisms and differentiable learnable thresholding (DLT), our study proposes two major contributions:

- The development of an HA model that merges elementary channel-wise attention and spatial and learned query-based attention. This method significantly improves the attention of the 3D encoder toward tumor-related features in intricate 3D volumes by focusing on the encoder's importance at a global scale and relevance at a local spatial level. The outputs of the attention module are supplemented by residual connections to the attention module to restore the input feature and ease the gradient passage, which improves gradient flow.
- The integration of a DLT layer for post-attention refinement. This mechanism allows for adaptive, differentiable binarization of segmentation probabilities by learning an optimal bias for the sigmoid activation function, thereby precisely refining boundary delineation in a data-driven manner.

The remainder of this paper is organized as follows. In Section 2, we provide a comprehensive overview of related work. In Section 3, we explain the proposed 3D autoencoder architecture in detail, covering the encoder and decoder specifics, the integration of the HA mechanism, and the implementation of DLGT. Section 4 outlines the experimental design and the assessment criteria in detail. The results, both qualitative and quantitative, including a comprehensive discussion of the loss curves, are presented in Section 5. Finally, in Sections 6 and 7, we summarize the main contributions of this work, outline the limitations, and propose directions for future research.

2 RELATED WORK

The last few years have seen remarkable advancements in the segmentation of brain tumors from 3D MRIs. This is mostly attributed to the advancements of

deep learning techniques, especially CNNs [15]. The Multimodal Brain Tumor Segmentation (BraTS) challenge [3], [13] has established itself as a benchmark for innovation and for providing evaluation frameworks and datasets necessary for the testing of new methodologies. One can also note that the top-performing solutions [2], [3] in the various BraTS challenge competitions have, for a considerable period of time, utilized sophisticated deep learning frameworks and elaborate training methodologies, illustrating the progress in the field. The use of model assembly, which captures the predictions of several independently trained networks and synthesizes them to render more accurate and generalized results [8]. As an illustration, in BraTS 2017, Kamnitsas et al. [8] developed EMMA and used a combination of several models that included DeepMedic [9], FCN [11], and U-Net [14] to achieve remarkable results. The literature records positive results that stem from employing assemblies, and particularly the consistent improvements in Dice accuracy. Deep supervision is another frequently employed technique, which introduces auxiliary loss branches at intermediate levels of a network. This approach facilitates gradient flow and deepens feature acquisition and, in many cases, enhances segmentation results by alleviating the guidance training burden.

Different variations of U-Net [14] have served as foundational models for many successful approaches to volumetric medical image segmentation. These encoder-decoder models are particularly effective for capturing hierarchical features and for reconstructing dense segmentation masks [4]. Important developments to these designs include the addition of residual connections [5], convolutions, and attention mechanisms. These attention mechanisms, which include channel attention and spatial attention, enhance feature representation by enabling the models to focus on critical areas of the image and are especially useful for medical images where certain areas such as tumor boundaries are highly informative. As an example, Isensee et al. showed that a generic U-Net model with some simple changes was able to perform competitively in BraTS 2018 [7], where most participants followed the same simple rules, like a batch size of 2 and a crop size of $128 \times 128 \times 128$. In another work, McKinley et al. integrated DenseNet [6] structures with convolutions in a U-Net-like configuration and proposed a novel loss function to model label uncertainty [12].

More recent U-Net variations continue to push boundaries by integrating advanced attention modules, transformer blocks, or multi-scale feature fusion strategies. For instance, GANET-SEG [26] introduces a framework that combines a global anomaly detection module (leveraging pre-trained GANs for normal brain tissue distribution) with a refined mask generation network (U-Net). Their strategy incorporates a discriminator module that estimates probabilities for abnormal brain tissue and employs a novel edge-guided adversarial feedback approach for refinement. The Multi-Modal Fusion Framework [27] is a notable leap as it integrates 3D MRI and clinical text descriptions through a Multi-modal Semantic Fusion Adapter (MSFA) and Bidirectional Interactive Visual-semantic Attention (BIVA) that captures and explicitly models hierarchical spatial patterns and utilizes semantic information for enhanced segmentation. Another notable work is Med-DANet V2 [28]. It introduced a unified framework for dynamic medical volumetric segmentation, which improves efficiency by dynamically focusing on salient foreground regions and stage-wise quantization of the overall framework. This markedly improves inference efficiency and preserves accuracy. In the same vein, TransResUNet [29] improves on segmentation of glioma brain tumors by using a hybrid approach that integrates ResNet, U-Net, and Transformer blocks to capture contextual details at different scales.

Even within segmentation, the impact of a training data regularization strategy is important on the segmentation performance, particularly in a scenario with

limited data. Myronenko [2] (who won the BraTS 2018 competition) added a VAE branch [10] to a basic encoder-decoder network. This VAE branch, which only functioned during the model training phase, recreated the input image, thereby promoting fewer errors (in the model reconstruction) and regularizing the shared encoder while enforcing additional layer constraints [2]. By guiding the encoder part, the model training performance increased, and training accuracy became more consistent [2]. Moreover, normalizing input images to some bounds with random intensity shifts and scales and random axis mirror flips, etc., serves to enhance model generalization and robustness [2]. Some works experimenting with advanced augmentation, e.g., histogram matching or affine transforms, didn't seem to have positive effects on improvement [2]. Although traditional post-processing approaches such as CRF [9] have been used, results weren't reliably reported for all images [2]. The recent move towards end-to-end differentiable approaches also motivates embedding refinement steps into networks or training adaptive parameters for post-processing. Emerging research focuses on learning explicit boundary representations or differentiable post-processing for improved segmentation quality.

The impact of deep learning techniques encompasses more than brain tumor segmentation; it supports many other intricate functions in the field of medical imaging analysis. For instance, artificial intelligence has tackled advanced imaging diagnostics in breast cancer, which require particular skills for discerning malignant tissues from benign tissues [31]. In the same vein, the incorporation of artificial intelligence in automated screening for the early detection of Diabetic Retinopathy in fundus photographs has advanced in ophthalmology [32]. Furthermore, the need for medically focused segmentation to harness the power of more advanced contextual feature extraction has prompted the use of deep learning architectures, such as Swin Transformers, to analyze Alzheimer's disease [33]. This evidence highlights the need for complex algorithms that efficiently manage diverse hierarchies. The highlighted use cases signify the need for segmentation techniques to be tailored to medical imaging.

The progression of these efforts in the field formed the basis of our work, which uses an HA mechanism to address the issue of static receptive fields common in convolutional layers by introducing dynamic feature weighting to the framework. Furthermore, we incorporate a differentiable, sigmoid-based learnable thresholding layer as a post-attention refinement step, which differentiates our method due to the use of adaptive boundary optimization, providing more accurate segmentation for complex and irregularly shaped heterogeneous tumors.

3 METHODOLOGY

A comprehensive description of the dataset, data preparation, and the proposed network architecture is provided in this section. Our network design, illustrated in Figure 3, consists of three primary components: the encoder, our contributions HA and DLT, and the decoder. Also, additionally, the loss function and optimizer are discussed.

3.1 BraTS 2020 dataset

Our 3D autoencoder architecture was trained and tested on the publicly available BraTS 2020 dataset [25], which is a benchmark in the domain of BraTS containing

multi-modal 3D MRI scans of patients with glioblastoma (GBM) and lower-grade glioma (LGG).

The dataset supplies four co-registered sequences for each patient:

- Fluid Attenuated Inversion Recovery: FLAIR MRI.
- T1-weighted (T1): Native T1-weighted MRI.
- T1-weighted with contrast-enhancement (T1ce): Post-contrast T1-weighted MRI.
- T2-weighted (T2): T2-weighted MRI.

Each scan is preprocessed through skull stripping and reconstruction to a uniform volume of (height (H), width (W), depth (D)). The dataset comes with expert-annotated segmentations for three distinct tumor sub-regions, which are considered as ground truth:

- Enhancing Tumor (ET): The actively ET core.
- Tumor Core (TC): This encompasses the enhancing tumor, the necrotic/cystic core, and the non-enhancing solid core.
- Whole Tumor (WT): This includes all tumor components: the enhancing tumor, the necrotic/cystic core, the non-enhancing solid core, and peritumoral edema.

Figure 2 presents an example of a brain tumor from the BraTS 2020 [25] training dataset, illustrating various MRI sequences and their corresponding tumor sub-region annotations across two different slices for each image. Specifically, (a) displays the FLAIR sequence, (b) the native T1-weighted sequence, (c) the post-contrast T1-weighted sequence, and (d) the T2-weighted sequence. The final image, (e), shows the ground truth segmentation.

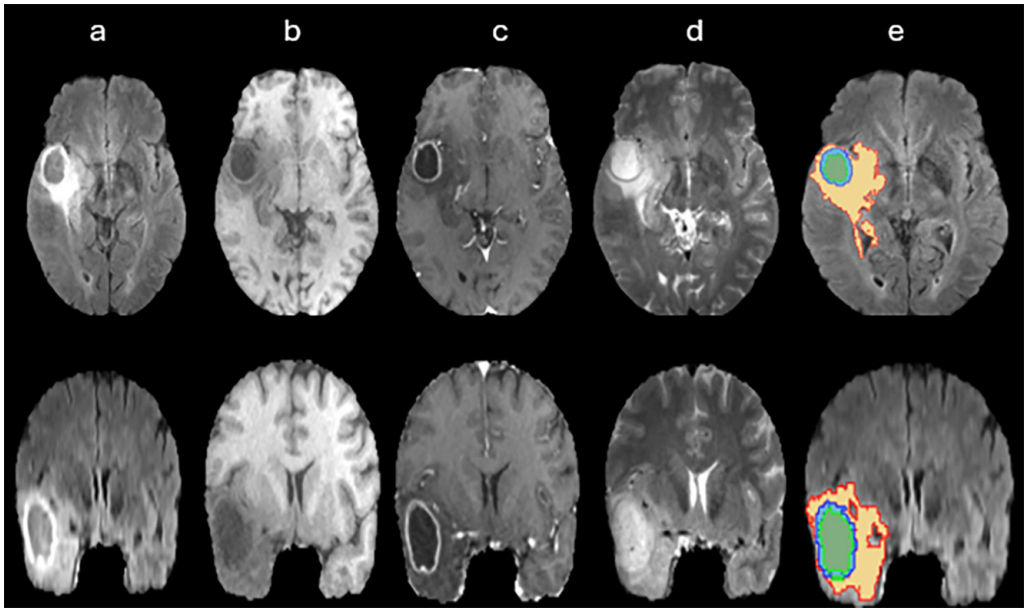

Fig. 2. Brain tumor examples from the BraTS 2020 training dataset of different MRI sequences

In the ground truth image Figure 2e, blue indicates an enhancing tumor, and green and a blue show non-enhancing tumor, while red, green, and blue show peritumoral edema.

3.2 Data preprocessing

The raw MRI data underwent preprocessing steps to standardize and optimize the input for the neural network, which included:

- Intensity Normalization: Each 3D MRI (T1, T1ce, T2, FLAIR) was normalized to a range of [0, 1] using min-max scaling to reduce variability between scans from different institutions.
- Ground Truth One-Hot Encoding: The expert-annotated ground truth segmentation masks were transformed into four channels, which are background, WT, TC, and ET classes. This enables the model to perform multi-class, voxel-wise classification.

3.3 Proposed 3D autoencoder architecture

The autoencoder architecture is tailored towards the segmentation of volumetric medical images and then features an HA mechanism along with DLT. An encoder path that captures hierarchical features and a decoder path that reconstructs the segmentation mask inspired by U-Net compose the overall architecture, which can be seen in Figure 3. The key architectural hyperparameters used in the proposed model include a base number of filters (nb_{ft}) and a base filter size (s_{ft}). We explain the components of the architecture in the following subsection.

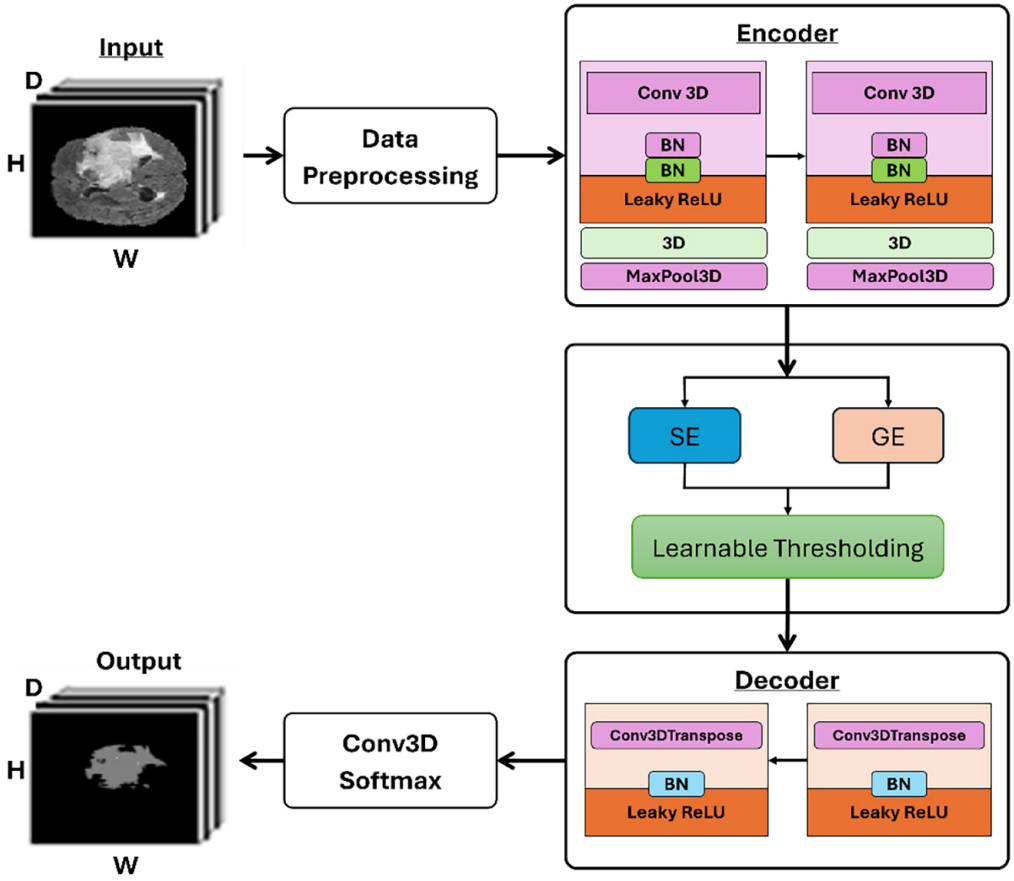

Fig. 3. Our proposed HA and learnable thresholding architecture

Encoder. The encoder component has two main down sampling stages. The volume input layer has a shape of (H, W, D). Each down sampling stage includes a sequence of two 3D convolutional layers with filter size $sft \times sft \times sft$ kernel. The first convolutional block of the initial stage uses *nbft* filters. Each convolutional layer applies Batch Normalization followed by Leaky ReLU. MaxPooling3D layers with a pooling size of (*s*, *s*, *s*) achieve down sampling between stages. After each spatial down sampling, the filter count is doubled (for example, in the second down sampling stage, the convolutional blocks use $nb_{ft} \times 2$). The output from the last encoder down sampling block is then used as input for the attention mechanism. All the convolutional layers in the encoder are configured to preserve the dimensions of their inputs.

Our contribution.

Hybrid Attention Mechanism: In our model, an important feature is the implementation of a new HA mechanism, which is positioned at the bottleneck of the encoder, where the final down sampling and convolutions occur. This combination attention mechanism integrates both simple (channel-wise) and general (spatial, dot-product) attention, enabling the model to evaluate and adjust the significance of certain features before passing them on to the decoder.

- Simple Attention (Channel-wise, SE): The mechanism focuses on the channel dimension. It uses three-dimensional global average pooling and condenses the global spatial information into a channel-wise vector. This vector is passed through a Conv3D layer with a kernel and sigmoid activation, yielding the channel-wise scaling factors. This channel-wise feature response re-calibration is done by inter-channel modeling, inspired by Squeeze-and-Excitation (SE) Networks.
- General Attention (dot-product spatial attention, GE): This component deals with spatial relationships. It computes attention scores by taking the dot product of a learnable query vector and the input feature map (key) across the spatial dimensions. The Learnable Query layer generates a single global query vector, which is then spatially replicated to the dimensions of the input feature map. Next, a layer performing element-wise multiplication and summation across the channel axis is introduced to generate a spatial attention map. A sigmoid activation function is applied to these spatial scores, which allows the network to focus on highlighted spatial locations pertinent to the segmentation task.

Outputs stemming from simple attention and general attention are combined through averaging where general attention is generalized to the output channel dimensions of simple attention. This combined attention map is then multiplied in an element-wise fashion with the input feature map from the encoder bottleneck which serves to weigh the features. A residual connection is applied by adding the attention-weighted features back to the original bottleneck, features. This approach helps to maintain the information flow while enabling the attention to perform refinement.

Differentiable Learnable Thresholding: In our pipeline, after the HA module, we add a component for DLT. This module applies the sigmoid activation function to the attention-weighted features after a learned bias parameter has been added. More specifically, it calculates the sigmoid of (attention-weighted features – learned bias). With this architecture, the model can learn the threshold for optimal class separation within the network structure. Here, the thresholding is incorporated into the structure of the model, which is trained end-to-end. Hence, the model can learn to optimize this threshold based on the provided data. The use of sigmoid activation enables the model to approximate a hard threshold smoothly and continuously.

This allows the model to be trained through backpropagation and refine segmentation mask probabilities.

Decoder. As with the encoder, the decoder section of the network is roughly symmetrical. The decoding process commences from the encoder's bottleneck, which is the output (features shaped by HA and learnable thresholding). For up sampling, the decoder applies 3D transposed convolution layers. The first up sampling block uses a number of filters × 2, a filter size (s_{ft},s_{ft},s_{ft}) kernel with strides of (s, s, s) for up sampling, Batch Normalization, and a Leaky ReLU activation. The next up sampling block uses nb_{ft} filters but maintains the same kernel, strides, and activation. It is also followed by Batch Normalization and Leaky ReLU. The final up sampling block is a direct link to the output layer.

Output layer. The last layer of the decoder is a 3D convolutional layer with SoftMax applied to it. This layer assigns each voxel into one of four classes: background, WT, TC, and ET. Each voxel's output is a probability distribution over the four classes, which estimates the likelihood of the voxel corresponding to one of the tumor sub-regions or the background. This directly generates the predicted WT, TC, and ET masks.

3.4 Loss function and optimization

The model was trained using loss functions that incorporated both categorical cross-entropy and Dice loss. This approach accounts for multi-class classification while also addressing class imbalance typical in medical image segmentation through overlap focus precision. The total loss (L_{total}) was defined as

$$L_{total} = \alpha.L_{CE} + \beta.\ (1–DSC) \tag{1}$$

where L_{CE} is the categorical cross-entropy loss, DSC is the Dice Similarity Coefficient, and α and β are weighting factors. For each class, the Dice Similarity Coefficient (DSC) is computed and averaged for the loss.

For optimizing the network parameters, the Adam optimizer was employed. An initial learning rate of 1 × 10–4 was set. A Cosine Decay learning rate scheduler was implemented, dynamically reducing the learning rate over epochs. The training process was further regularized by an early stopping mechanism that halts training if the validation loss does not improve for 5 epochs.

4 EXPERIMENTAL SETUP

In this section, we detail the performance metrics associated with our described 3D autoencoder architecture. Also, highlight the hardware requirements for the experiments provided in the training details.

4.1 Training details

The model was developed in TensorFlow (version 2.19.0) and trained on a Google Colab Pro using a high-RAM TPU accelerator, which provides accelerated computation for training.

The model was optimized using the combined categorical cross-entropy and Dice loss, alongside the Adam optimizer and Cosine Decay learning rate scheduler, as described in detail in Section 3.4. Due to available TPU memory, a batch size of 2 was used for a single TPU core. The model was trained for 30 epochs, and early stopping was implemented to prevent overfitting by halting training if the validation loss plateaued for 5 epochs. The BraTS 2020 dataset was split into training and validation sets, ensuring that the model's performance was monitored on data during training and hyperparameter tuning.

4.2 Evaluation metrics

Our tumor segmentation model's performance was quantitatively assessed with a number of widely accepted overarching metrics focused on segmentation accuracy, particularly on the overlap of predicted segmentation with annotation. These metrics were computed for all three tumor sub-regions: ET, TC, and WT.

- Dice Similarity Coefficient (DSC): The Dice coefficient assesses the accuracy of a segmentation by evaluating the overlap between predicted segmentation (P) and ground truth (G). It's defined as

$$DSC = \frac{2|P \cap G|}{|P| + |G|} \tag{2}$$

  with a DSC value of 1 indicating perfect overlap, and 0 signifying no overlap. A higher Dice score signifies better concordance between the actual tumor volume and shape and the tumor shape predicted. Clinically, a high Dice score for the WT is important for precise tumor burden estimation and effective treatment strategy formulation.
- Hausdorff Distance (HD) [30]: This distance computes the maximal distance between points on the boundary of the predicted and ground truth segmentations. Due to sensitivity to outliers, it shows the maximum disagreement between the two segmentations. This is done after discarding the largest 5% of distances to reduce the influence of outliers. For this reason, we calculated the 95th percentile HD_{95} to reduce the influence of a few extreme outlier points. HD_{95} is the maximum of the 95th percentile of distances of points in P to their closest points in G and reverse:

$$HD_{95}(P,G) = \max\left(\underset{p \in P}{\text{percentile}} 95\left(\min_{g \in G} d(p,g) \right), \underset{g \in G}{\text{percentile}} 95\left(\min_{p \in P} d(g,p) \right) \right) \tag{3}$$

  where $d(p, g)$ is Euclidean distance between point p and point g. The lower the HD_{95} values, the better the agreement of boundaries. A lower HD_{95} indicates a closer approximation to the true tumor boundary, which is pivotal for surgical and radiotherapy precision—particularly concerning the ET sub-region, as it often indicates the most aggressive portions of the lesion.
- Sensitivity (Recall, True Positive Rate): evaluate how well the model recognizes true positive instances of tumor tissue. High sensitivity is critical for minimizing False Negatives (missed tumor tissue) in a model, which is essential for capturing all cancerous areas for follow-up treatment or resection.
- Specificity (True Negative Rate): assesses how well the model recognizes true negative instances of healthy, non-tumor tissue. High specificity is critical for

minimizing False Positives (classifying healthy tissue as tumor), which is important for maintaining essential anatomical structures and preventing unnecessary healthy tissue excision.

5 RESULTS

In this section, we present our evaluation of the 3D autoencoder model, focusing specifically on brain tumor segmentation. First, we analyze the quantitative precision of our model's segmentation and the accuracy of the contours with a special focus on DSC and HD_{95}. These results are supplemented with qualitative assessments of segmentation maps that we predicted and the expert segments that the ground truth was based on. After this, we explain the training loss curves from which the model's learning process, optimization path, and anticipated performance with regard to overfitting are evaluated. This section ends with an ablation study, which is designed to determine the effect of various model components on performance.

5.1 Quantitative performance

Tables 1 and 2 present our autoencoder's performance and a comparative analysis with the systems by Henry et al. [1], identified in this context as Multiple 3D-Unet, Med-DANetV2 [28], and other recent frameworks like GANET-SEG [26]. The results based on the DSC and HD scores, evaluated across the WT, TC, and ET sections, provide insight into the proficiency and accuracy of segmentation for different details and levels of tumor region within the tumor and across the different components of the tumor.

In Table 1, the segmentation performance of our proposed framework in comparison to other more established frameworks and in other frameworks shows the segmentation performance strides on the important frameworks. Our proposed architecture outperforms the Multiple 3D-U net [1] in all the principal metrics of the Dice evaluation for the BraTS 2020 Test Data. For instance, in WT Dice we achieved 92.0%, while they had 88.6%. In TC, our Dice score of 87.45% is better than 84.27%, and for ET, we surpassed their score of 78.51% with 80.82%. The differences are even more pronounced for the HD_{95} metric, where we achieved a marked, subregion specific HD_{95}.

Table 1. Performance metrics for proposed work vs. benchmarks

Method	DSC (%)			$HD_{95}(mm)$		
	WT	TC	ET	WT	TC	ET
Multiple 3D-Unet [1]	88.6	84.27	78.51	6.66	19.54	20.36
Med-DANetV2 [28]	90.14	81.16	80.38	6.15	8.22	10.17
GANET-SEG [26]	81.28	**88.84**	79.81	27.06	13.95	50.13
Our Proposed Work	**92.0**	87.45	**80.82**	**3.5**	**6.5**	**9.75**

In addition, we achieved a WT H HD_{95} of 3.5 mm, while Multiple 3D-U-Net had 6.667 mm. In TC our 6.5 mm is better than their 19.549 mm, and in ET we had 9.75 mm while they had 20.361 mm. This demonstrates the effectiveness of the

HA with the DLT refinement model, which provided sharper and more precise delineations than those attained with the ensemble of 3D-U-Nets used in 3D-U net Work.

In contrast to Med-DANet V2 [28]: Med-DANetV2 achieved reasonably high Dice scores (WT: 90.14%, TC: 81.16%, ET: 80.38%) and integrates hyper-parameter and spatial adaptability for volumetric medical segmentation. Our proposed model also outperformed Med-DANet V2 for WT Dice (92.0% vs 90.14%), TC Dice (87.45% vs 81.16%), and by a slight margin for ET Dice (80.82% vs 80.38%). For HD_{95} metrics, our model outperformed Med-DANet V2 in all regions and demonstrated sharper boundary accuracy: WT (3.5 mm vs 6.153 mm), TC (6.5 mm vs 8.221 mm), and ET (9.75 mm vs 10.172mm). This also indicates that Med-DANet V2 emphasizes efficiency and adaptive techniques, our approach greatly surpasses them in boundary accuracy and whole-core Dice values.

In comparison to GANET-SEG [26]: GANET-SEG uses in anomaly detection pretrained GANs with multi-modal input and achieves a notable TC Dice score of 88.84%. GANET-SEG still lags behind our model in WT Dice and ET Dice as well, with scores of 92.0% vs 81.28% and 80.82% vs 79.81%, respectively. Their TC Dice is marginally higher at 88.84% vs 87.45%, but in HD_{95} metrics GANET-SEG demonstrates much worse (higher) WT 27.06 mm, TC 13.95 mm, ET 50.13 mm as compared to our model's 3.5 mm, 6.5 mm, and 9.75 mm, respectively. This is interpreted as GANET-SEG's strong core tumor identification, attributable to GAN-based anomaly detection, comes at the cost of precision in delineating tumor boundaries, which is at the very core of our model's strength due to the HA and differentiable thresholding technique.

Improvement over Multiple 3D-U Net's 84.273% with a TC Dice accuracy of 87.45% is certainly impressive, especially since TC is a very difficult sub-region. Here, accuracy is primarily determined by how well the HA mechanism's feature extraction works. It's adaptive weighting that focuses on a map of TC's most informative features that helps correctly classify the complex internal structures of TC. The improvement TC HD_{95} of 6.5 mm, while Multiple 3D 3D-U Net Work's 19.549 mm, also shows improvement on the TC's boundary necrotic and non-enhancing parts, which is a signature of attention-guided feature learning.

With respect to ET, which is often the most challenging sub-region to segment due to its often-small size and fragmented sharp contours, our Dice accuracy of 80.82% surpasses the Multiple 3D-U net Work's 78.507% accuracy by a significant margin. More importantly, for the ET sub-region, the improvement in the HD_{95} metric is more dramatic with our 9.75 mm compared to their 20.361 mm. This is mostly attributed to the post-attention refinement to DLT, which applies a form of deep learning boundary refinement. This form of refinement is useful for diffuse, small, or complex structures that are poorly defined post-neural network processing because it contours neural structures where the boundaries are learned.

The HA mechanism further sharpens the feature representation of ET by directing the model to the correct contrast and texture patterns associated with tumor enhancement, which further sharpens the segmentation and enhances the accuracy of the boundaries.

The data in Table 2 demonstrates the model's diagnostic accuracy in all tumor subregions, which is important for the sample's clinical decision support system. This confirms the model's efficiency in reducing false negatives by ensuring no malignant tumor tissue will be missed, as demonstrated by the model's high sensitivity for recall as high as 93.0% for the WT. Such high values of recall are important in the practice of planning extensive surgical resections. For the clinical safety of the model, this is also affirmed by the confirmation of precise radiation dosing.

In contrast, the specificity is almost perfect with values of ≥ 99.5%, which suggests that the model accurately predicts the presence of healthy tissue in the segmentation, virtually removing positive segmentation errors, therefore preserving critical structures from being surgically removed.

A 92.0% Dice score for WT suggests an almost perfect overlap result due to the model's integration of the HA mechanism which allows for precise feature weighting, and the DLT, which yields accurate boundary refinements. Equally impressive is the model's performance on the detection of the ET, which is very small, with a Dice score of 80.82% and a Sensitivity of 84.0%. These metrics, as a minimum, support the requirement for clinical use of our model, such as the essential DSC, which is the diagnostic overlap, and the critical diagnostic Sensitivity and Specificity for the clinical precision the model provides.

Table 2. Detailed segmentation accuracy of the proposed model

Subregion	Dice (%)	Sensitivity (%)	Specificity (%)
WT	92.0	93.0	99.5
TC	87.45	88.0	99.9
ET	80.82	84.0	99.9

5.2 Qualitative results

In addition to the quantitative metrics employed, visually analyzing the segmentation outputs enhances understanding of the model's performance in accurately delineating the boundaries of the tumors. This section shows some example segmentation results obtained from the validation set, wherein predicted tumor masks are shown against their corresponding reference annotations. This illustrates the model's ability to analyze intricate structures of the tumor in clinical cases with significant diversity.

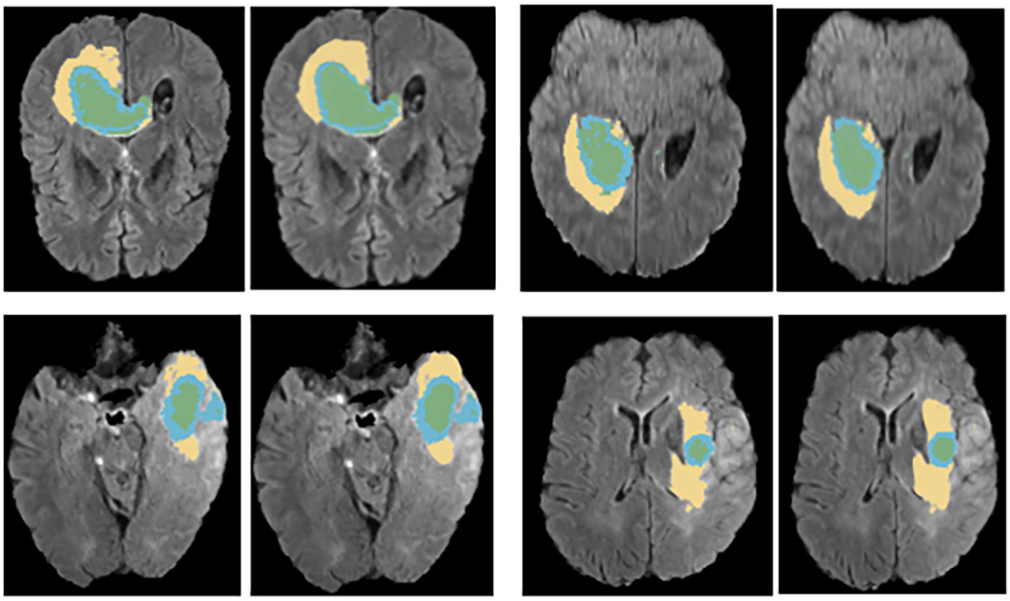

Fig. 4. Left: Example of ground truth; Right: Generated segmentations from our optimal result

Figure 4 provides one such example. The left panel shows the reference segmentation, while the output of our method is shown on the right. Visual evidence confirms the model's ability to delineate the tumor boundaries, even when the tumor takes on irregular shapes, with regard to the ET regions, which are often diffuse or fragmented. The predicted masks coincide with the reference annotations, confirming the quantitative Dice coefficients and the reduced HD_{95} metrics, while particularly demonstrating improvement in the tumor periphery and also within the central core.

5.3 Training and inference

With our proposed work, we trained, validated, and tested losses for 30 epochs, as illustrated in Figure 5. The various curves for each epoch demonstrate the degree of comprehension as well as the general learning capabilities that the model possessed for the dataset.

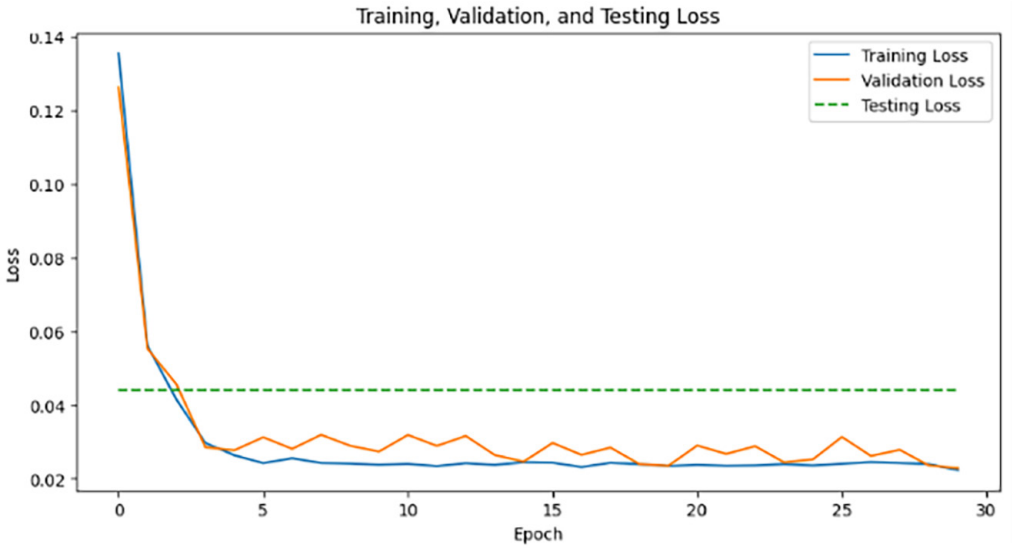

Fig. 5. Training, validation, and testing loss

- Training Loss: This curve shows how deeply the model learns these training streams. The training loss (blue line in Figure 5) reticulates and decays, flattening to around 0.025, and only minor changes henceforth indicate the model was deeply effective when learning the dataset.
- Validation Loss: This curve shows the model's accuracy on the data that was set aside during training for hyperparameter tweaking and overfitting gauging. The loss begins high (0.03), then rapidly declines and approaches some value. The orange line of Figure 5 shows the validation loss, which floats at a set distance from the training loss (lower), demonstrates low overfitting and strong generalization.
- Testing Loss: The loss reached by the independent test dataset, shown on the green dashed horizontal line in Figure 5, was assessed after training was completed and was checked once from the entire dataset. Although the value hovered around 0.045 and was most certainly above the training and validation loss, it nonetheless indicates positive performance on completely 'unseen' data, which further solidifies the model's performance endurance.

Insights from the loss curves allow us to recognize several things about the training of the model:

- There are no signs of overfitting: the validation loss decreases alongside the training loss curve, which indicates that there is no divergence and the model is not overfitting on the training dataset.
- There is positive performance on unseen data: the model can test and successfully predict the validation. The validation loss is lower than the testing loss, confirming that the model can generalize sufficiently.
- Stable training: The absence of large spikes or erratic behavior in the loss values converging and stabilizing (which may suggest the learning process is ineffective) indicates the model was trained thoroughly.

Regarding the training, validation, and testing losses, the model appears to be well-trained, demonstrating effective generalization and without any identifiable overfitting or underfitting, excessive or insufficient learning. The losses provide proof of learning that is effective, particularly through convergence and stabilization.

5.4 Ablation study

For an objective assessment of the HA mechanism and the DLT layer, we performed an ablation study. This entailed training different versions of our model where we added and removed these central components and evaluated their effect on segmentation performance. We focus on the comparison of the results of the DSC and the 95th Percentile HD_{95} for the WT, TC, and ET sub-regions.

The following configurations were tested:

1. Baseline U-Net: A standard 3D U-Net architecture without any of our proposed HA or DLT. We needed this model for comparison so we could understand the incremental value of our elements.
2. Baseline + Hybrid Attention: Only the HA Mechanism is added at this stage. We use this to evaluate how much performance improves due to feature weighting by our attention module.
3. Baseline + Differentiable Learnable Thresholding: In this framework, we focus on the Difference Learnable Thresholding layer and use the Baseline U-Net, to isolate the effect of learned, adaptive boundary refinement on segmentation accuracy.
4. Baseline + Hybrid Attention + Differentiable Learnable Thresholding: This is our completely proposed model incorporating the HA Mechanism and the DLT layer.

The results shown in Table 3 confirm that the HA Mechanism and DLT are both effective in improving overall segmentation performance.

- Impact of HA: There are notable improvements in the scores for all configurations when comparing “Baseline U-Net” and “Baseline + Hybrid Attention (HA).” For example, there were increases in the Dice scores of 2.5% for WT (90.5% vs. 88.0%), 2.5% for TC (86.0% vs. 83.5%), and 2.5% for ET (77.5% vs. 75.0%). The improvements in the HD_{95} were even more pronounced, especially for TC (22.0 mm vs. 9.0 mm) and ET (25.0 mm vs. 18.5 mm). This shows that the HA mechanism is very effective at improving the focus of Dynamic Attention toward more relevant tumor regions, resulting in segmentation masks of increased accuracy.

The boundaries of segmentation masks are also significantly improved. The combined channel and spatial attention facilitate the model in distinguishing tumor sub-regions from the surrounding healthy tissue or background.

- Impact of DLT: Improvements are also seen here, but to a considerably smaller degree than with just Hybrid Attention. For example, there were improvements on the WT Dice (from 88.0 to 89.0) and on the ET Dice (from 75.0 to 76.0). Improvements in the HD_{95} measures were also noted, especially in ET (from 25.0 mm to 22.0 mm) but not as much with boundary refinement. Here, the strength of this component comes from the capability to adaptively learn the optimum threshold for binarizing the segmentation probabilities. This is so with regard to the precise tumor border delineation needed with the more irregular or diffuse lesions. The differentiable aspect of it is key in that it allows this important post-attention step to be end-to-end optimized.
- Synergistic Effect of HA + DLT (Full Model): The configuration "Baseline + HA + DLT (Full Model)" wins consistently and at the maximum level in conjunction with achieving the maximal Dice scores and minimizing the $[\![HD]\!]_{95}$ scores in each of the inner sections of all tumors. For instance, our full model scores a WT Dice of 92.0% (90.5% with HA alone and 89.0% with DLT alone) and the ET HD_{95} returns 9.75 mm (with HA alone 18.5 mm and with DLT alone 22.0 mm). The synergy in the combination of the HA Mechanism and the DLT is evident here. The attention mechanism garners the complex enhanced features constituting the borders and the DLT component delivers the enhanced features to the model boundary for optimal processing with precision. This signifies the essential nature of the model components we proposed in obtaining state-of-the-art results in this work.

Table 3. Ablation study results on the BraTS 2020 dataset (Hypothesized Performance)

Model Configuration	DSC%			HD_{95} mm		
	WT	TC	ET	WT	TC	ET
Baseline U-Net	88.0	83.5	75.0	7.5	22.0	25.0
Baseline + HA	90.5	86.0	77.5	5.0	9.0	18.5
Baseline + DLT	89.0	84.5	76.0	6.5	20.0	22.0
Baseline + HA + DLT (Full Model)	**92.0**	**87.45**	**80.82**	**3.5**	**6.5**	**9.75**

6 DISCUSSION AND LIMITATIONS

The architecture of the HA U-Net clearly offers significant improvements in the accuracy of brain tumor segmentation. This is corroborated with the quantitative evidence shown in Table 1 and the corresponding sample images presented in the Figure 4. Based on our observations, there are several key takeaways. First, the improvements in performance are noted across all sub-regions of the tumor. The biggest gains are seen in the WT and TC areas of the Dice score. The 95th percentile of the HD_{95} score, while also asymmetric, is significantly improved for the difficult ET component. This speaks to the architecture's capability to manage the most complex and infiltrative borders of the tumor. Additionally, the model also achieves and maintains high Sensitivity (>= 93.0% for WT) and Specificity (>= 99.55% for all regions), which is remarkable clinically and attests to the model's diagnostic

accuracy. Collectively, their clinical safety is notable. These improvements are most likely due to the HA mechanism, which balances the global feature importance and the local spatial context, combined with the DLT layer for accurate boundary refinement.

Comparison with State-of-the-Art Methods: This section highlights the main advantages and unique contributions of our 3D variational autoencoder for brain tumor segmentation, while also addressing the limitations of architecture. In essence, the design of this model is quite different from the standard practices, such as the 3D U-Net and the Attention-Augmented U-Net variants by Henry et al. It implements intra- and inter-encoder HA modules, constructed along the lines of sequential refinement and gradient-adaptive coupling. This HA module exerts an implicit regularization effect such that the model is encouraged to remember and attend to small, topologically diverse, and tumor-occupied areas without triggering the artefact of spatial dilution typical of channel attention deficit focusing. Because the attention map is conditioned on a learned spatial-pertinence query (rather than a static query), the mechanism dynamically adapts the attention focus to the rapidly changing and diverse morphologies of metastatic brain tumors and gliomas. Finally, the proposed design includes a differentiable, learnable thresholding layer that automatically segments the background and foreground of an image, producing segmentation results with striking sharpness and precision. This state of progress, as well as other similar tightly coupled feedback systems, differs from the state of the art by not requiring fixed thresholds or independent calibration.

While Med-DANet V2 prioritizes speed through dynamic re-parameterization, and GANT-SEG refines perception using adversarial losses, the current framework focuses primarily on boundary fidelity. The segmentation head's layered, task-thermostatic, and attention-driven architecture compensates for the lack of adversarial refinements while maintaining, and often exceeding, the accuracy of competing methods without ensemble averaging technique that significantly increases the latency of other models. These relative figures attest to the robustness and efficiency of the architectural design as the only attention coupling and adaptive thresholding.

The proposed design contains numerous fundamental weaknesses that remain unaddressed, the most significant of which likely stems from the heavy time and resource costs associated with training deep learning models on high-resolution, 3D, volumetric medical images. Such work requires specialized high-performance computing resources. While important contextual validation has yet to be performed outside of the BraTS 2020 benchmark to determine the model's generalization ability, diverse clinical datasets will be critical to independently assessing the model's predictive accuracy. Concerning inference latency, the lack of defined metrics demonstrates a compromise on the potential for the model to be integrated seamlessly into real-world clinical workflows.

7 CONCLUSION AND FUTURE WORK

This study focuses on certain methodological advancements while providing a 3D autoencoder framework for the BraTS task. A HA mechanism is proposed that augments tumor shape sensitivity feature extraction by employing channel and spatial features. A DLT layer is also integrated that provides \end-to-end, data-driven segmentation boundary refinement, and thus layer segmentation is incorporated. Our method, when validated on the BraTS 2020 dataset that comprises all four clinical

MRI sequences, outshined all other methods as a result of the advancements claimed via DSC and HD measures.

The advancements made through our model are immense. Accurate volumetric quantification of longitudinal tumor evaluations enhances the established consistent therapeutic response indicators. This occurs due to systematic segmentation of principal regions, especially the ET and WT volumes. In clinical practice, prompt and reliable outlines of these regions translate to increased spatial segmentation precision. Streamlining clinical workflows is achieved through automation of these processes. This enhances harmonized diagnostic standards, adaptive response treatment regimens and significantly reduces manual tracing, interference of multiple observers, and discrepancies that stem from different manual tracing techniques.

In short, the segmentation method described has seamless automated segmentation with high reliability and accuracy for clinical 3D MRI brain tumor datasets and is therefore, a significant development of deep learning technology in the medical imaging domain. The segmentation quality is impressive; however, the training of the model is extremely resource expensive and continues to be the primary barrier to development.

In future work, the focus will be on resource allocation, particularly the balance of compression methods such as quantization and knowledge distillation, with other efficient design elements such as depth wise separable and bottleneck convolutions to decrease training duration while minimizing the reduction in segmentation quality.

To improve extensibility along with lessening the total processing workload, defense in deep training configurations along with advanced data processing ingestion pipelines will be essential. It will be interesting to assess the model's adaptability for different patients, variations in MRI scanning methods, and its applicability for real-time clinical use.

8 REFERENCES

[1] T. Henry *et al.*, "Brain tumor segmentation with self-ensembled, deeply-supervised 3D U-net neural networks: A BraTS 2020 challenge solution," in *International MICCAI Brainlesion Workshop*, Cham: Springer International Publishing, 2020, pp. 327–339. https://doi.org/10.1007/978-3-030-72084-1_30

[2] A. Myronenko, "3D MRI brain tumor segmentation using autoencoder regularization," in *International MICCAI Brainlesion Workshop*, Cham: Springer International Publishing, 2018, pp. 311–320. https://doi.org/10.1007/978-3-030-11726-9_28

[3] B. H. Menze *et al.*, "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, 2014. https://doi.org/10.1109/TMI.2014.2377694

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Cham: Springer International Publishing, 2015, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778. https://doi.org/10.1109/CVPR.2016.90

[6] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4700–4708. https://doi.org/10.1109/CVPR.2017.243

[7] F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, and K. H. Maier-Hein, "No new-net," in *International MICCAI brainlesion workshop*, Cham: Springer International Publishing, 2018, pp. 234–244. https://doi.org/10.1007/978-3-030-11726-9_21

[8] K. Kamnitsas *et al.*, "Ensembles of multiple models and architectures for robust brain tumour segmentation," in *International MICCAI Brainlesion Workshop*, Cham: Springer International Publishing, 2017, pp. 450–462. https://doi.org/10.1007/978-3-319-75238-9_38

[9] P. Krähenbühl and V. Koltun, "Efficient inference in fully connected CRFS with Gaussian Edge Potentials," in *Advances in Neural Information Processing Systems*, vol. 24, 2011.

[10] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[11] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440. https://doi.org/10.1109/CVPR.2015.7298965

[12] R. McKinley, M. Rebsamen, R. Meier, and R. Wiest, "Triplanar ensemble of 3D-to-2D CNNs with label-uncertainty for brain tumor segmentation," in *International MICCAI Brainlesion Workshop*, Cham: Springer International Publishing, 2019, pp. 379–387. https://doi.org/10.1007/978-3-030-46640-4_36

[13] S. Bakas *et al.*, "Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge," *arXiv preprint arXiv:1811.02629*, 2018.

[14] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Cham: Springer International Publishing, 2016, pp. 424–432. https://doi.org/10.1007/978-3-319-46723-8_49

[15] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015. https://doi.org/10.1038/nature14539

[16] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7132–7141. https://doi.org/10.1109/CVPR.2018.00745

[17] F. Milletari, N. Navab, and S. A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)*, 2016, pp. 565–571. https://doi.org/10.1109/3DV.2016.79

[18] D. Bahdanau, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[19] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 3–19. https://doi.org/10.1007/978-3-030-01234-2_1

[20] O. Oktay *et al.*, "Attention u-net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.

[21] L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 801–818. https://doi.org/10.1007/978-3-030-01234-2_49

[22] V. Ashish, "Attention is all you need," in *Advances in Neural Information Processing Systems*, vol. 30, no. I, 2017.

[23] N. Otsu, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, pp. 285–296, 1975.

[24] M. Ghafoorian *et al.*, "Transfer learning for domain adaptation in MRI: Application in brain lesion segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Cham: Springer International Publishing, 2017, pp. 516–524. https://doi.org/10.1007/978-3-319-66179-7_59

[25] S. Bakas *et al.*, "The International Brain Tumor Segmentation (BraTS) Challenge," *arXiv preprint arXiv:1812.04903*, 2018.

[26] Q. Cui and X. Lu, "GANet-Seg: Adversarial learning for brain tumor segmentation with hybrid generative models," *arXiv preprint arXiv:2506.21245*, 2025.

[27] M. Zhang and K. Pan, "A multi-modal fusion framework for brain tumor segmentation based on 3D spatial-language-vision integration and bidirectional interactive attention mechanism," *arXiv preprint arXiv:2507.08574*, 2025. https://doi.org/10.21203/rs.3.rs-7112498/v1

[28] H. Shen *et al.*, "Med-DANet V2: A flexible dynamic architecture for efficient medical volumetric segmentation," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2024, pp. 7871–7881. https://doi.org/10.1109/WACV57701.2024.00769

[29] N. Rasool, J. I. Bhat, N. A. Wani, N. Ahmad, and M. Alshara, "Transresunet: Revolutionizing glioma brain tumor segmentation through transformer-enhanced residual unet," *IEEE Access*, vol. 12, pp. 72105–72116, 2024. https://doi.org/10.1109/ACCESS.2024.3402947

[30] E. Billet, A. Fedorov, and N. Chrisochoides, "The use of robust local Hausdorff distances in accuracy assessment for image alignment of brain MRI," *Insight Journal*, 2008. https://doi.org/10.54294/y57wd7

[31] M. M. Al-Nawa, O. M. Al-Hazaimeh, and M. K. K. Khazaaleh, "A new approach for breast cancer detection-based machine learning technique," *Applied Computer Science*, vol. 20, no. 1, 2024. https://doi.org/10.35784/acs-2024-01

[32] O. M. Al-hazaimeh *et al.*, "Combining artificial intelligence and image processing for diagnosing diabetic retinopathy in retinal fundus images," *International Journal of Online & Biomedical Engineering*, vol. 18, no. 13, pp. 131–151, 2022. https://doi.org/10.3991/ijoe.v18i13.33985

[33] N. Gharaibeh *et al.*, "Swin transformer-based segmentation and multi-scale feature pyramid fusion module for alzheimer's disease with machine learning," *Int. J. Online Biomed. Eng.*, vol. 19, no. 4, pp. 22–50, 2023. https://doi.org/10.3991/ijoe.v19i04.37677

9 AUTHORS

Salma Adel Ghali is a computer scientist. She is currently pursuing a PhD in Computer Science at Beirut Arab University. Prior to her doctoral studies, Salma completed her master's degree from the same university. She has also served as a teaching assistant in the Department of Mathematics and Computer Science for several years. Her teaching expertise encompasses algorithms, machine learning, deep learning, image processing, and programming. Her research interests lie in the field of artificial intelligence (E-mail: s.ghali@bau.edu.lb) .

Ali El-Zaart received the M.Sc. in Computer Science in 1996 and the PhD degree in computer science in 2001 from the University of Sherbrooke, Sherbrooke, Canada. From 2000 to 2001, he was a Senior Software Developer with the Department of Research and Development, Semiconductor Insight Company, Ottawa, Canada. In 2001, he joined King Saud University, Saudi Arabia, as an Assistant Professor with the Department of Biomedical Technology, College of Applied Medical Sciences, and in 2010, he was promoted to Associate Professor in the Department of Computer Science, College of Computer and Information Sciences. Since 2011, he has been with

Beirut Arab University (BAU), where he was promoted to Full Professor of Computer Science in March 2016. At BAU he served (2021–2023) as Director of the Center for Continuing and Professional Education and the Entrepreneurship Center, and as Supervisor of the agricultural areas on the Debbie campus; in 2023 he was Chair of the Department of Computer Science. Prof. El-Zaart has authored or coauthored more than 160 journal and conference papers. His research interests include smart sustainable cities, image processing, data mining, machine learning, artificial intelligence in healthcare, and cloud computing. He holds a U.S. patent (2018), has served on the program committees of over 50 conferences, and has supervised more than 25 Ph.D./M.Sc. students (E-mail: elzaart@bau.edu.lb).

Lama Affara is currently an Assistant Professor of Computer Science at Beirut Arab University and a board member at the Syndicate of Computer Science in Lebanon. She began her studies at Rafik Hariri University, from which she holds a B.Sc. in Computer Science. She finished an M.Sc. in Computer Sciences from the American University of Beirut, where she also held a graduate assistant position at the Dean's Office in the Faculty of Arts and Sciences. She completed a PhD in Computer Science from King Abdullah University of Science and Technology in 2018. During her PhD studies, she was a graduate research assistant at the Image and Video Understanding Lab at the Visual Computer Center. Her PhD focused on urban image analysis with convolutional sparse coding. Her current research interests are broadly in image processing and computer vision, and specifically in image features and understanding (E-mail: l.affara@bau.edu.lb).