Healthcare Analytics 5 (2024) 100307



Contents lists available at ScienceDirect

Healthcare Analytics

journal homepage: www.elsevier.com/locate/health





A triplanar ensemble model for brain tumor segmentation with volumetric multiparametric magnetic resonance images

Snehal Rajput [a], Rupal Kapdi [b], Mohendra Roy [a,*], Mehul S. Raval [c,*]

[a] School of Technology, Pandit Deendayal Energy University, Gandhinagar, 382007, Gujarat, India
[b] Institute of Technology, Nirma University, Ahmedabad, 382481, Gujarat, India
[c] School of Engineering and Applied Science, Ahmedabad University, Ahmedabad, 380009, Gujarat, India

ARTICLE INFO

Keywords:
Automated segmentation
Brain tumor
Ensemble
Machine learning
Medical image analysis

ABSTRACT

Automated segmentation methods can produce faster segmentation of tumors in medical images, aiding medical professionals in diagnosis and treatment plans. A 3D U-Net method excels in this task but has high computational costs due to large model parameters, which limits their application under resource constraints. This study targets an optimized triplanar (2.5D) model ensemble to generate accurate segmentation with fewer parameters. The proposed triplanar model uses spatial and channel attention mechanisms and information from multiple orthogonal planar views to predict segmentation labels. In particular, we studied the optimum filter size to improve the accuracy without increasing the network complexity. The model generated output is further post-processed to fine-tune the segmentation results. The Dice similarity coefficients (Dice-score) of the Brain Tumor Segmentation (BraTS) 2020 training set for enhancing tumor (ET), whole tumor (WT), and tumor core (TC) are 0.736, 0.896, and 0.841, whereas, for the validation set, they are 0.713, 0.873, and 0.778, respectively. The proposed base model has only 10.25 M parameters, three times less than BraTS 2020's best-performing model (ET 0.798, WT 0.912, TC 0.857) on the validation set. The proposed ensemble model has 93.5 M parameters, 1.6 times less than the top-ranked model and two times less than the third-ranked model (ET 0.793, WT 0.911, TC 0.853 on validation set) of BraTS2020 challenge.

1. Introduction

Glioblastoma is the most malignant and frequently occurring primary tumor in the central nervous system. In the United States, 25% of newly diagnosed primary brain tumor cases are glioblastoma [1]. The median survival rate is about ten months despite significant technological advances [2]. There is no known cure for this disease, and the current clinical practices of a safe gross total resection (GTR) through surgery followed by chemotherapy and radiation therapy are widely accepted [1]. Diagnosing glioblastoma is challenging due to the high heterogeneity in lesions, shapes, and locations. Automated segmentation can help oncologists diagnose and plan treatment, increasing patients' survival tenure.

MRI is widely used because it is non-invasive and provides high resolution and contrast for the soft tissues, which helps segment highly heterogeneous tumor lesions [2–4]. However, due to image artifacts and the complexity of tumors, a single anatomical MRI is not enough to segment all tumors. For various tumors, multiparametric MRI (mpMRI) provides supplementary information, and it includes four MR pulse sequences: T1-weighted (T1), T1-enhanced, T2-weighted (T2), and T2-weighted with fluid-attenuated inversion recovery (FLAIR). These MR pulse sequences are regarded as reliable tools for detecting enhancing tumor (ET)/necrosis (NC) and peritumoral edema (ED), respectively.

Recently, many methods have been proposed for 3D image construction [5], preprocessing [6], network engineering [7], and postprocessing [7], which collectively focus on improving brain tumor segmentation. Researchers have introduced enhanced versions of the 2D and 3D-based U-Net architecture in the context of brain tumor segmentation (BTS). These variants incorporate advancements such as residual connections [8–10], dense connections [11–13], attention mechanisms [14–17], and self-attention mechanisms like transformers [18–20], among other techniques. Though these models have increased segmentation accuracy, their complexity and computation requirements are very high. In practical applications, networks with robust performance, low computation need, and possessing simple structure are necessary [21,22]. In clinical usage, achieving the best accuracy while operating within limited computational budgets is challenging.

* Corresponding author.
E-mail addresses: mohendra.roy@ieee.org (M. Roy), mehul.raval@ahduni.edu.in (M.S. Raval).

https://doi.org/10.1016/j.health.2024.100307
Received 3 September 2023; Received in revised form 4 January 2024; Accepted 29 January 2024
Available online 4 February 2024
2772-4425/© 2024 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Along with 3D, 2D models are also prevalent for BTS. Also, most sizeable open-source datasets are two-dimensional images, which implicitly forces 2D processing [23,24]. The significant difference between 3D and 2D networks is that 3D networks capture spatial and temporal relationships within the data, leveraging the additional depth dimension for feature extraction, whereas 2D networks capture only spatial dependencies within the images. These additional depth dimensions in 3D models helped segmentation improvement but with higher computational costs. Moreover, the 3D models do not guarantee better tumor segmentation and require a large dataset, whose preparation is costly and time-consuming. Also, 3D models are difficult to optimize and require more training time. In comparison, 2D models do not have access to depth information, which can limit their performance but requires lesser computations. Therefore, the triplanar model was introduced to capitalize on the advantages of both 3D and 2D models [25].

The triplanar approach utilizes three 2D models, each specifically trained to analyze the axial, sagittal, and coronal planes. This enables the networks to capture unique patterns from various perspectives, leading to improved segmentation and smoother continuity across slices compared to singular 2D models [25,26]. When predicting results, segmentation maps from these three models are combined, enabling the model to integrate multiplanar information, yielding better segmentation. Moreover, this triplanar model balances computations similar to 2D convolutional neural networks (CNN) and can capture some depth information. This model can effectively handle anisotropic data (pixel size varies across planes due to differences in image acquisition protocols) by independently processing each plane, adapting to variations in resolution, and producing accurate segmentation [23]. Furthermore, utilizing and optimizing a triplanar network has the potential to produce robust segmentation results that can compete with contemporary segmentation models [27,28]. Therefore, in our work, we leverage the triplanar model and attention mechanism for BTS. The attention mechanism focuses on specific parts of an input, guided by their corresponding weights, which reflect the importance of the information they carry. Higher weights signify more valuable information within a specific feature map, enabling the model to downplay irrelevant aspects. In the channel attention mechanism, weight is assigned to each channel with respect to the relevance of information. On the other hand, the spatial attention mechanism assigns weight to the prominent characteristics within the feature map, confined by the kernel's window. However, in standard convolutional layers, all channels are given equal weight, which can lead to redundancy, inefficiencies in feature extraction, and slow convergence. The advantages of these attention networks lie in their adaptive fine-tuning of weights based on information, making them well-suited for capturing intricate semantic relationships, including small-intensity dissimilarities, and preserving long-term dependency [17,29]. It has a simple architecture, suitable to plug and use. Additionally, they offer interpretability, a crucial feature in medical image segmentation.

This study focuses on the benefits of triplanar networks and attention mechanisms, particularly in scenarios with restricted memory and computation environments. The proposed work involves creating an ensemble of these networks with an attention mechanism. The objective is to enhance the model's performance without causing parameter explosion. For this purpose, we conduct ablation studies focusing on 2D U-Net-based Triplanar models and attention mechanisms to improve the model's segmentation capabilities. The ablation study examines the impact of different attention mechanisms: channel attention (*CA*) [30], concurrent combining of channel and spatial attention (*CCSA*) [31], and sequential (*SCSA*) [32]. However, we found the *CCSA* model's performance lags behind CA and *SCSA*, prompting us to optimize *CCSA* (called *CCSAv1*) by widening the filters at each layer to enhance its expressiveness and discriminative ability [31,32]. This *CCSAv1* network has shown notable performance enhancements among all the planar models. Lastly, we employ post-processing techniques to improve further the segmentation results. It is important to note that the architectures of spatial and channel attention differ between *CCSA* and *SCSA*.

The remaining sections of the paper are organized as follows: Section 2 provides an overview of the methodology and data sources. Section 3 presents the results and discussions. Section 4 concludes the proposed research work, addresses the limitations and presents future work.

2. Methodology and data

The methodology includes details of end-to-end procedures used to generate robust segmentation results while ensuring that the network has a simple structure and uses limited memory and computation requirements. This is essential for wide-scale and real-world applications in the medical domain [22,33,34]. The holistic structure of the proposed methodology is illustrated in Fig. 1. It is a triplanar ensemble-based approach where three orthogonal planar models (axial, sagittal, and coronal) were trained on all three combinations of attention-based models *CA*, *CCSAv1* and *SCSA* respectively. The ablation study also aims to examine the impact of CA, concurrently combining *CCSA* or sequentially combining *SCSA* channel and spatial attention mechanisms to improve the results. This combining approach enables the model to comprehend 'what' to focus on (channel information) and 'where' to focus on (spatial information) within the feature map. This allows the model to better understand the salient features in the data across spatial and channel dimensions, which can enhance the discriminating abilities of the model [31,32]. Further, we observed that *CSCAv1* network has shown notable performance enhancements among all the orthogonal planar models. Finally, we obtained the segmentation results by averaging the probability maps of the ensemble. Subsequently, we created a *Super-ensemble* by combining all nine models and post-processed the segmentation results using the connected component algorithm to derive the final segmentation. We kept only the bigger connected component in the predicted segmentation mask. Every combination of such planar and Triplanar models can be seen in Fig. 1.

2.1. Data sources and scoring metrics

The proposed network was trained using the training dataset of the mpMRI Brain Tumor Segmentation Challenge (BraTS) 2020, which consists of 3D images [35–37]. The training dataset comprises 369 samples collected from 19 institutions, each with distinct image-acquisition protocols and scanners. The ground labels were manually annotated by one to four experts. The details of mpMRI are shown in Table 1. The validation and test dataset includes 125 and 166 samples, respectively. The ground truth of the MRI images is labeled with specific values to represent different regions of interest: the necrotic center and non-enhancing tumor were labeled as 1, the edema region as 2, the enhancing tumor region as 4, and the background as 0. The ground truth of validation and testing sets was withheld by the challenge organizers' datasets, which will be used later to rank the segmentation network. However, test sets were available only for the challenge participants. Each modality has a volume measurement of 240 (width) $\times$ 240 (height) $\times$ 155 (depth or slices). The challenge organizer has done preprocessing, such as skull-stripping and interpolating to a voxel resolution of $1 \times 1 \times 1$.

The Dice-score and Hausdorff distance (HD) are mainly used as scoring metrics in the BraTS2020 challenge. These metrics are utilized to evaluate the effectiveness of the segmentation algorithms. The Dice-score measures the spatial overlap between the predicted segmentation labels (P_s) and the ground-truth labels (G_t), and its value ranges from 0 to 1, where a score of 0 indicates no overlap, and a score of 1 indicates a complete match between the predicted and ground-truth labels. A high Dice-score indicates that the segmentation results closely

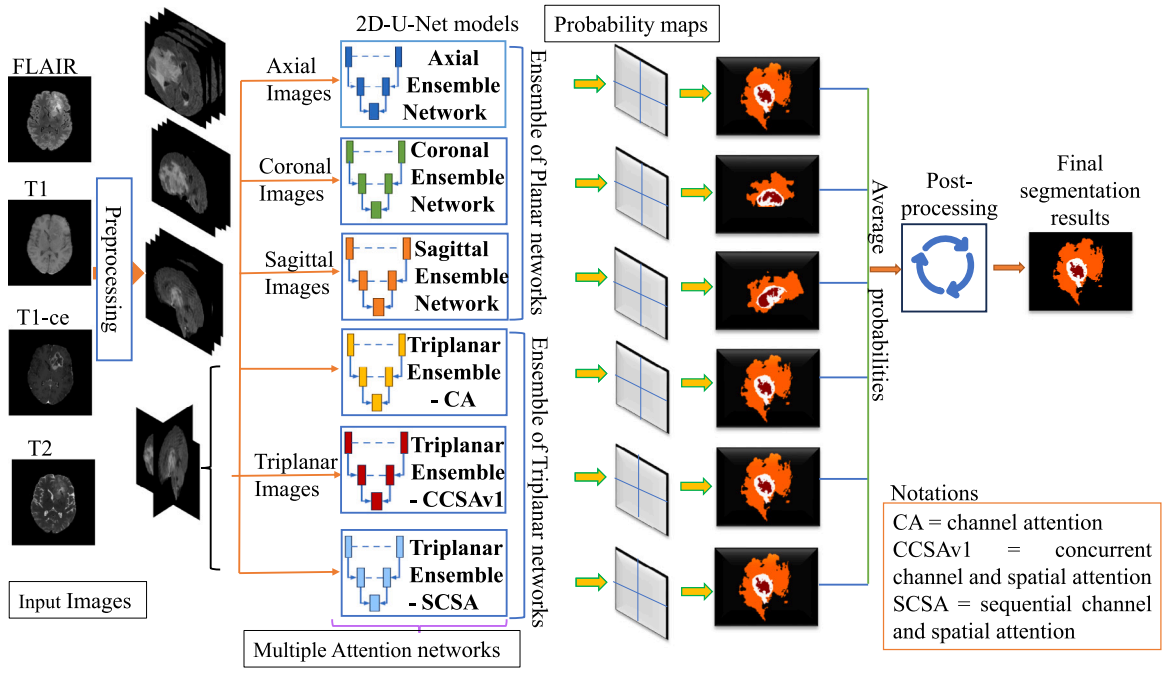


Fig. 1. The holistic structure of the proposed methodology.

Table 1
Details about the mpMRI dataset [37].

Pulse Sequences	Acquisition type (2D/3D)	Slice thickness/Voxel size
T1	2D - sagittal or axial	1 – 6 mm
T1-ce	3D	1 mm^3
FLAIR	2D - axial, sagittal or coronal	2 – 6 mm
T2	2D - axial	2 – 6 mm

align with expert annotations. On the other hand, the HD evaluates the alignment of the predicted segmentation boundary with the ground-truth boundary. It measures the greatest of all distances between a point in one set (G_t) and the closest point in the other set (P_s). Smaller HD signifies better-aligned boundaries between G_t and P_s, whereas larger HD signifies poor alignments. The Dice-score is sensitive to false positives and negatives, emphasizing the overall match between the segmented areas. Meanwhile, the HD focuses on the maximum deviation, highlighting the most significant differences between the segmentations. By considering both metrics, researchers and practitioners can gain a more nuanced understanding of the segmentation quality, which is crucial for diagnosis, treatment planning, and monitoring of diseases. Mathematically, Dice-score and HD can be defined as:

$$Dice - score = \frac{2 \cdot |P_s \wedge G_t|}{(|P_s| + |G_t|)/2} \tag{1}$$

$$HD = \max\left\{ \underset{g \epsilon G_t}{S} \underset{p \epsilon P_s}{I} d(g,p), \underset{p \epsilon P_s}{S} \underset{g \epsilon G_t}{I} d(g,p) \right\} \tag{2}$$

Where S and I represent supremum, infimum, respectively, and $d(g,p)$ is the distance between points $g \epsilon G_t$ and points $p \epsilon P_s$. Also, alternative metrics like sensitivity and specificity are employed by the BraTS challenge to assess potential over-segmentation or under-segmentation by the algorithms.

2.2. Data preprocessing

Since MRI scans frequently exhibit uneven intensity distribution due to bias fields, we initially corrected bias using the N4ITK package [38]. Additionally, non-brain areas were eliminated from all the MRI MR pulse sequences. Subsequently, the top and bottom 1% of intensities, considered outliers, were removed. Finally, intensity normalization using mean and standard deviation values was conducted on each slice of the images. The dimensions of the input image slices are 192 × 152 for an axial model, 192 × 144 for the sagittal model, and 152 × 144 for the coronal model. The slices have different sizes due to the orientation of the imaging plane and the way the MRI is processed. While training, we randomly flipped images vertically and horizontally on the fly to reduce memory usage and provide better training.

2.3. Network design and training details

The primary network consists of a layered encoder–decoder architecture, as shown in Fig. 2. The encoder pathway captures features ranging from low-level to increasingly high-level as it processes the input images. At each consecutive stage, the width of the channels doubles, and the spatial size of the feature maps reduces by half. In the first stage, the encoder consists of a 2 × 2 convolution layer followed by a residual block. All other stages consist of a downsampling layer followed by a residual block. The residual block has a ResNet-like structure comprising two 3 × 3 kernel size convolution blocks, batch-normalization, and Parametric ReLU (PRelu) unit [39]. For downsampling, convolution with stride length 2 is used. At the same time, the decoder is primarily responsible for capturing the precise location of features in the image. It works with the encoder to reconstruct the segmentation mask by upsampling the feature maps to the original input size. Here, the decoder consists of an upsampling layer followed by 2 × 2 convolution layer, residual block, and attention block [30,32]. Finally, softmax activation was applied to generate four feature maps, one for each class.

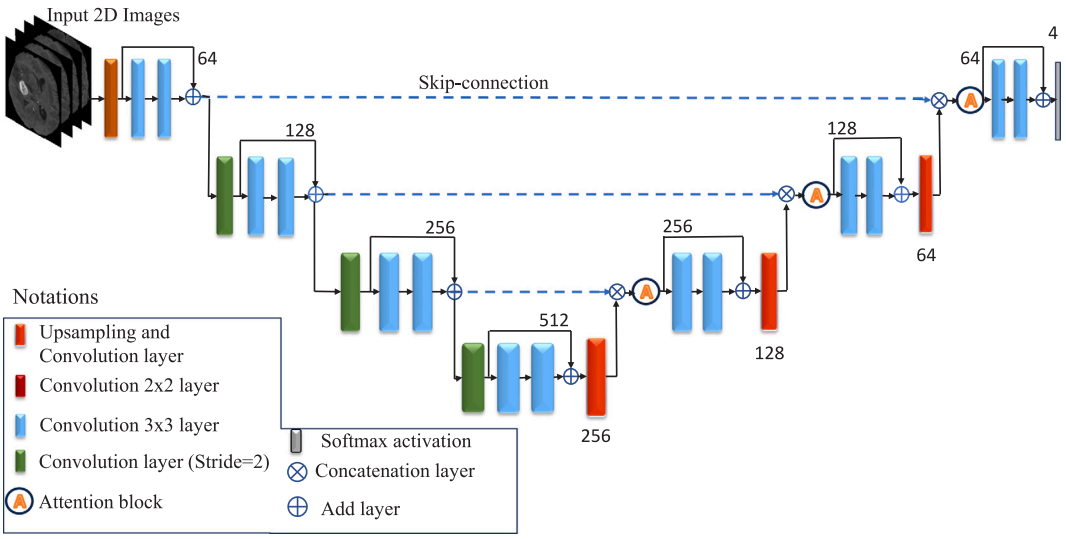


Fig. 2. The architectural diagram of the proposed basic 2D U-Net with built-in attention mechanism at the decoder [40].

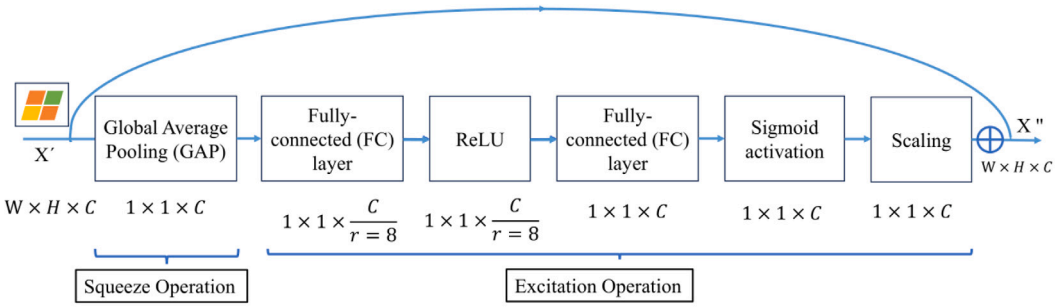


Fig. 3. The modified *CA* mechanism's with *ratio* value is 8 [30,40].

2.3.1. Rearrangement of the attention architecture

We employed the squeeze and excitation (SE) technique [30] to implement CA, which can be seen in Fig. 3. The SE consists of two key operations, namely squeeze and excitation. The squeeze operation involves the global average pooling (GAP) method to squeeze the spatial dimension of the feature map into a vector. The excitation operation follows the squeeze operation and aims to understand dependencies and relationships between channels. It involves two fully connected layers, typically with a non-linear activation function, which helps the network learn to emphasize informative features while suppressing less useful ones. As shown in Fig. 3, the first fully connected (FC) layer decreased the channel size (C) to $C/ratio$. While the native method [30] recommended using $ratio = 16$. Empirically, we found that $ratio = 8$ yielded the best results for the BraTS 2020 dataset. Adopting $ratio = 16$ would significantly reduce the number of channels, limiting the model's ability to capture context effectively.

Similarly, the structure of the improved *CCSAv1* mechanism is illustrated in Fig. 4, demonstrating the parallel combination of channel attention and spatial attention. For channel attention, we have employed SE. As for spatial attention, the convolution layer with a 1×1 kernel performs the squeeze operation, followed by an excitation operation where a sigmoid function is applied to generate weights multiplied by the original feature map. During experimentation, we noted that (*CCSA*) is not performing well compared to the other CA and SCSA attention mechanisms. Therefore, we modified Roy et al. [31] Spatial and Channel Squeeze and Excitation Block (scSE) module in *CCSA*, and called as *CCSAv1*. In the native method, [31], squeezing spatial attention involves using $C \times K$, where C represents the number of filters, and K denotes the 1×1 kernel size. In the native method, with each depth of U-Net, C remains constant at 1, whereas in our improved *CCSAv1* mechanism, C increases with the respective depth of the U-Net, starting with an initial value of 64 and reaching up to 512. The improved *CCSAv1* keeps more information by compressing it to a slightly larger size and retains more critical details without excessively inflating the model size. Adjusting the number of filters with the depth of the U-Net allows the model to adapt and learn representations better suited for the given complexities, potentially resulting in improved understanding and segmentation. Furthermore, this dynamic adjustment of the number of filters based on the U-Net's depth allows the model to adapt to varying scales of features and objects within the images, contributing to better overall segmentation performance.

The *SCSA* module can be seen in Fig. 5. Here, the channel attention mechanism is implemented using the *CA* module, where instead of GAP, two separate modules (each for max pooling and average pooling) with the same architecture were built and aggregated to weigh the feature map. The first FC layer's output channel (C) was also reduced to $C/ratio$. Here, also empirically, we found that $ratio = 8$ is generating the best results. The internal operations of *SCSA* can be seen in Fig. 5a. Mathematically, it can be defined as:

$$F_c' = A_c(F_{map}) \oplus F_{map} \tag{3}$$

$$F_s'' = A_s(F_c') \oplus F_c' \tag{4}$$

where, $\oplus$ indicates element-wise multiplication, F_{map} is a feature map, $A_c(F_{map})$ is channel attention map, $A_s(F_c')$ is spatial attention map and F_s'' is the final output.

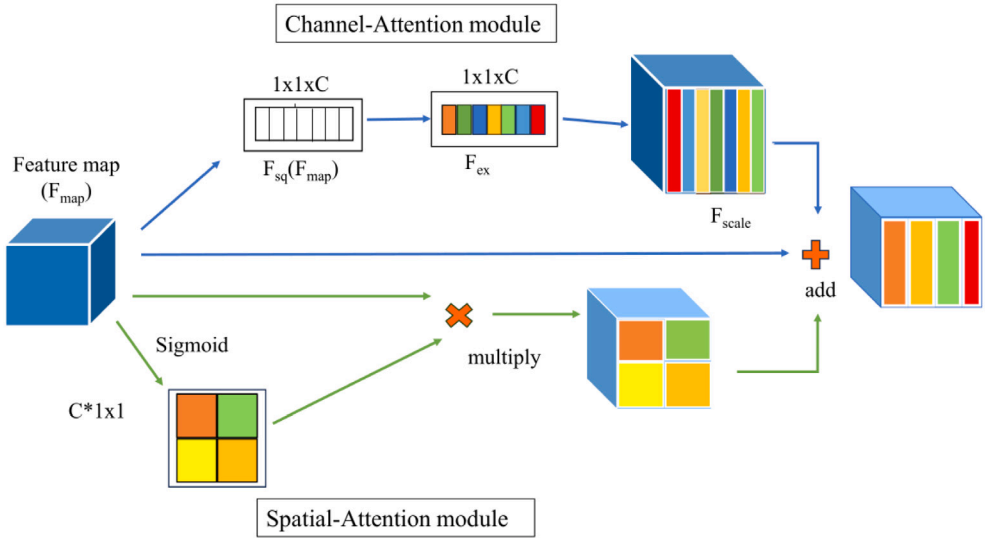


Fig. 4. The improved *CCSAv1* mechanism [31].

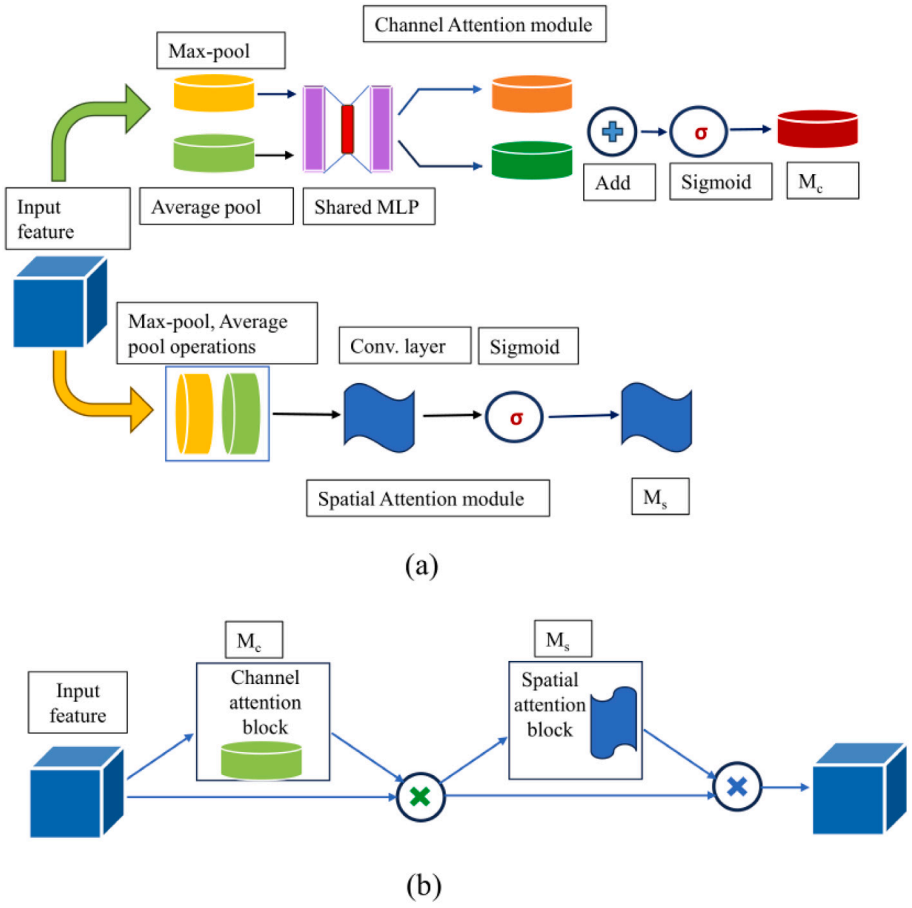


Fig. 5. The modified *SCSA* module with ratio = 8. (a) The working principle (b) overview of *SCSA* module [32,40].

2.4. Training details

The segmentation network was built using Tensorflow and Keras on a Quadro RTX5000 with 16 GB GPU and 128 GB RAM. During training, random slices of size 192 × 152 for an axial model, 192 × 144 for the sagittal model, and 152 × 144 for the coronal model were generated from the input volumes, and the batch size is set at 16 to accommodate our memory constraints. A stochastic gradient descent optimizer was used to train the network, and its learning rate was set to 0.008. The combination of cross-entropy (CE_{loss}) and generalized dice-loss functions (GD_{loss}) [41] was utilized as a loss function to tackle the class imbalance, as demonstrated in Eq. (5). The learning rate will decrease by 0.5 if the validation loss does not decrease for 30 epochs. Furthermore, the Keras early stopping mechanism was

Table 2
Experimental evaluation of each variant of models on BraTS2020 validation dataset. The best results are highlighted in bold text. Training and validation results from the same model are highlighted using the same color, and their corresponding training results can be found in Table 5. The two ranking models of the BraTS2020 challenge are shown in the first two rows. For the *CCSA* and *CCSAv1* of the same model type, matching colors are used (excluding the first two rows). All the results shown here are obtained from BraTS-challenge online assessment platform+.

Dataset	Model Type	Model Name	Parameter Numbers in millions (M)	Dice-score			HD		
				ET	WT	TC	ET	WT	TC
Validation	Ensemble of five 3DU-Nets	Isensee et al. [7]	30.2 × 5 = 156 M	0.798	0.912	0.857	26.410	3.730	5.640
Validation	Ensemble of eleven 3DU-Nets	Yuan et al. [15]	16.5 × 11 = 181.5 M	0.793	0.911	0.853	18.196	4.097	5.888
Validation	Axial	CA	10.246 M	0.610	0.873	0.754	56.627	9.532	14.038
Validation	Axial	CCSA	10.247 M	0.618	0.854	0.738	52.666	13.759	17.896
Validation	Axial*	CCSAv1*	10.589 M	0.679	0.867	0.788	42.230	10.885	11.496
Validation	Axial	SCSA	10.333 M	0.670	0.866	0.760	46.227	9.328	15.22
Validation	Coronal	CA	10.246 M	0.633	0.836	0.747	54.469	19.712	22.547
Validation	Coronal	CCSA	10.247 M	0.629	0.827	0.743	58.344	23.447	22.538
Validation	Coronal*	CCSAv1*	10.589 M	0.662	0.840	0.740	42.864	16.491	25.476
Validation	Coronal	SCSA	10.333 M	0.642	0.825	0.738	52.114	25.978	19.806
Validation	Sagittal	CA	10.246 M	0.663	0.850	0.737	48.509	10.414	15.969
Validation	Sagittal	CCSA	10.247 M	0.652	0.835	0.759	49.567	14.381	15.010
Validation	Sagittal*	CCSAv1*	10.589 M	0.663	0.833	0.747	48.829	10.955	18.954
Validation	Sagittal	SCSA	10.333 M	0.680	0.853	0.756	46.402	17.448	18.758
Validation	Axial Ensemble	CA-CCSA-SCSA	30.825 M	0.661	0.879	0.767	41.232	6.754	12.034
Validation	Axial Ensemble*	CA-CCSAv1-SCSA*	31.167 M	0.683	0.883	0.784	42.620	6.285	11.863
Validation	Coronal Ensemble	CA-CCSA-SCSA	30.825 M	0.648	0.849	0.762	53.748	16.502	15.713
Validation	Coronal Ensemble*	CA-CCSAv1-SCSA*	31.167 M	0.664	0.850	0.759	47.828	14.301	18.964
Validation	Sagittal Ensemble	CA-CCSA-SCSA	30.825 M	0.682	0.863	0.762	43.714	11.391	16.799
Validation	Sagittal Ensemble*	CA-CCSAv1-SCSA*	31.167 M	0.681	0.857	0.758	48.279	10.795	14.991
Validation	Triplanar Ensemble	CA	30.737 M	0.669	0.871	0.771	44.512	6.929	12.558
Validation	Triplanar Ensemble	CCSA	30.739 M	0.667	0.860	0.773	45.395	9.764	13.708
Validation	Triplanar Ensemble*	CCSAv1*	31.767 M	0.699	0.863	0.777	35.713	7.173	14.529
Validation	Triplanar Ensemble	SCSA	30.999 M	0.693	0.868	0.776	42.274	10.120	13.989
Validation	Super-Ensemble	CA-CCSA-SCSA	92.47 M	0.699	0.875	0.782	36.752	8.037	14.846
Validation	Super-Ensemble (without postprocessing)*	CA-CCSAv1-SCSA*	93.50 M	0.704	0.875	**0.785**	35.440	6.962	14.560
Validation	Super-Ensemble-pp (with postprocessing)*	CA-CCSAv1-SCSA*	93.50 M	**0.713**	**0.873**	0.778	**31.563**	**6.347**	13.274
Validation	Native U-Net *	without attention mechanism*	7.698 M	0.669	0.862	0.786	44.046	11.750	12.841

+ https://ipp.cbica.upenn.edu/
* signifies a modified version of models/

employed, which halts training if the validation loss does not decrease for 50 epochs.

$$\begin{aligned} Loss &= GD_{loss}(G_t, P_s) + CE_{loss}(G_t, P_s) \\ GD_{loss}(G_t, P_s) &= 1 - 2\frac{\sum_{l=1}^{C}(W_l \times \sum_{i=1}^{N} g_{li} \times p_{li})}{\sum_{l=1}^{C}(W_l \times \sum_{i=1}^{N} g_{li} \times p_{li})} \\ CE_{loss}(G_t, P_s) &= -\frac{1}{N}\sum_{i=1}^{N}\sum_{l=1}^{C}(g_{li} \times \log p_{li}) \end{aligned} \tag{5}$$

where $W_l = \frac{1}{(\sum_{i=1}^{N} g_{li})^2}$ is the adaptive weight for the lth channel, p_{li} is the predicted probability of class l for voxel i, is the ground truth label for class l for voxel i, $g \epsilon G_t$ where G_t is the ground truth label and $p \epsilon P_s$ where P_s is the predicted segmentation label.

3. Results and discussions

Table 2 shows the results of variants of attention mechanism trained on various planer images. In this context, we prioritize the discussion regarding *CCSAv1*, its parameter counts, its influence on segmentation as an individual model, and its potential ensembling with other model variants. We have also presented post-processed segmentation results. Additionally, we compared the proposed models with the network without attention mechanisms.

3.1. Impact of attention mechanism

The last row of Table 2 presents the results of the Native U-Net model, i.e., without attention mechanism, which is trained exclusively on axial planar images. It has 7.698 M parameters. While the proposed *CCSAv1* network has only a slightly higher number of parameters (an increase of 2.89 M) than Native U-Net. It demonstrates a notable improvement over the Native U-Net model of approximately 0.044 in Dice score for the ET region (where the Dice-score varies between 0 to 1), known to be a most challenging area for accurate segmentation. Similarly, for HD, it improved by 10 mm to Native U-Net.

3.2. Impact of improved $CCSAv1$ mechanism

The study of Table 2 reveals that the performance of the *CCSA* model was comparatively inferior to the *CA* and *SCSA* models, even when incorporating spatial and channel attention. For example, in Table 2, $4th\ row$ shows *CCSA* model information trained on the axial plane, which has Dice-score of 0.618 (ET), 0.854 (WT), 0.738 (TC) whereas *CA* model ($3rd\ row$) has 0.610 (ET), 0.873 (WT) 0.754 (TC) and *SCSA* ($6th\ row$) has 0.670 (ET), 0.866 (WT) 0.760 (TC) respectively. Similar observations can be obtained on sagittal (11th, 12th and $14th\ rows$) and coronal planes (7th, 8th and $10th\ rows$). The possible reasons for the poor performance of *CCSA* can be attributed to a simultaneous combination of the attention mechanism, which may not be integrating well with the U-Net architecture, it may need specific adjustments of the hyperparameters like a filter size, or it is less sensitive to the BraTS2020 dataset.

We modified *CCSA*'s spatial attention mechanism to increase the number of feature maps at each respective layer to obtain *CCSAv1*. This modification has increased the feature information and the network's expressive power with marginal addition to the network parameters [31]. In Table 2, a comparison between *CCSA* and *CCSAv1*

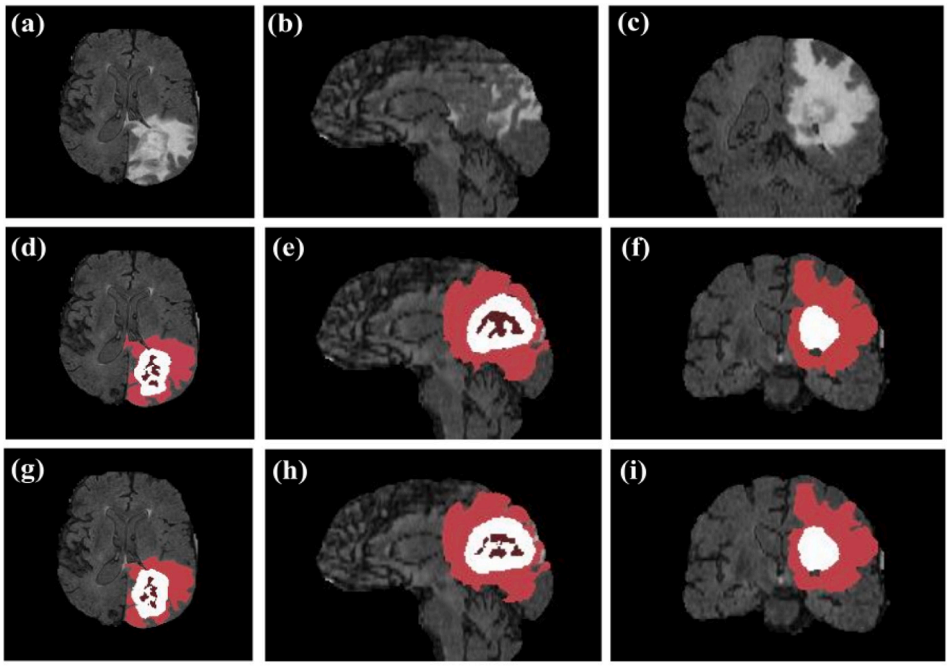


Fig. 6. (a), (b), and (c) are the axial, sagittal, and coronal views of the input FLAIR image. Whereas (d), (e), and (f) are the respective ground segmentation maps. Similarly, (g), (h), and (i) are the predicted segmentation-map for all the orthogonal plans view. Here, the WT is light-brown+white+dark-brown, the ET region is white, and the TC is dark-brown+white. Observing slices shown in (a), (b), and (c), we can easily infer that the axial slice contained the most detailed anatomical information, whereas the coronal had the least.

Table 3
Comparative analysis of computation and memory requirements between proposed models and Native U-Net.

	CA	*CCSAv1*	*SCSA*	*Native U-Net*
Trainable parameters	10.24 M	10.589 M	10.333 M	7.69 M
Training time per epoch	1290 ± 5 s	1302 ± 5 s	1302 ± 5 s	1260 ± 5 s
Memory size	15680 MB	15684 MB	15691 MB	15676 MB
Inference time per sample	1 s	1 s	1 s	1 s

on different model types is shown with the same color (excluding the first two rows). We can easily deduce that *CCSAv1* model has demonstrated noteworthy improvements compared to the other attention models (with all the combinations of orthogonal planar and triplanar models) with a marginal increase in the number of parameters. It has merely 0.342 *M* more parameters than the previous version *CCSA*. This validates that *CCSAv1* is a robust model.

The *Triplanar Ensemble* of *CCSAv1*, which is an ensemble of axial, sagittal, and coronal models having *CCSAv1* attention mechanism has achieved significant performance. The performance enhancement in the *CCSAv1* model is attributed to its triplanar nature, which provides access to depth information. The model type and name in red signify our approach's top-performing models. Further, *super ensemble model* with *(CA, SCSA, CCSAv1)* attention blocks have Dice-score 0.704, 0.875, 0.785 for ET, WT, and TC with total parameters 93.50 *M* which is 1.7× and 2× lesser than top ranking models shown in first two rows of the Table 2.

Further, postprocessing the segmentation results using the connected components algorithm helped remove false positive labels, which enhanced the Dice-score by 0.009 for the ET lesion segmentation (Dice-score (ET:0.713)). This can also be validated by looking at HD distance, which measures the distance between the predicted and segmented regions, which is the least in *Super ensemble-pp** i.e., model after post-processing (pp) shown in the bold font. Although *CCSAv1* as an individual model has contributed significantly in all the combinations, the *Super-ensemble model** (with *CCSAv1*),[1] when compared to *Super-ensemble model* (with *CCSA*) has only improved marginally for ET and Dice-score remained the same on WT and TC. This reason can be that averaging the probabilities obtained from multiple models will tone down significant changes from a particular model.

The comparative analysis of computation and memory requirements between proposed models and Native U-Net are shown in Table 3. We can observe marginal differences among the proposed methods and between the proposed method and the Native U-Net. Despite a slight increase in memory and computation requirements of the proposed triplanar model, segmentation performance has improved significantly.

3.3. Impact of imaging plane

Furthermore, comparing models based on the imaging planes, we can observe coronal-planar models performed inferior to other orthogonal (axial and sagittal) planar models. The significant evidence supporting this claim can be observed from Table 1, where it is apparent that the majority of MR pulse sequences are predominantly acquired in 2D, specifically in the axial or sagittal planes. Therefore, the coronal model lacks sufficient data for generalization, and hence, it has reduced coronal models' effectiveness. Similar observation can also be made by visualizing input FLAIR image in Fig. 6(a), (b), and

[1] * signifies a modified version of models.

Table 4
The average Dice-score of each orthogonal planar model. It is calculated by adding the Dice-scores of all three attention models.

Model	ET	WT	TC
Coronal	0.646	0.834	0.742
Sagittal	0.669	0.845	0.746
Axial	0.653	0.869	0.767

(c), where coronal views have less resolution than axial and sagittal; it is called anisotropy problem [23]. It occurs due to hardware limitations of MRI scan device, resulting in high resolution in the in-plane (axial) while exhibiting low resolution in the through-plane (coronal and/or sagittal) directions [42]. It creates unclear and not-detailed anatomical structure information, which hampers image visualization, analysis, and segmentation of lesions [42]. This can also be validated through the Dice score shown in Table 2. E.g., the average Dice-score of all the coronal, sagittal, and axial models are shown in Table 4. It shows that coronal models have the least average Dice-score.

3.4. Observations of segmentation on training set

Training results of top-performing models - triplanar and *Super-ensemble* on 75 randomly picked samples are presented in Table 5. Here also, we can observe that Triplanar ensemble models with $\approx 31M$ parameters can predict accurate labels for segmentation, having a Dice-score of 0.727, 0.892, and 0.829 for ET, WT, and TC. Whereas the *Super-ensemble* model with *CCSAv1* has shown improvement over all the lesions segmentation. Further, post-processing the segmentation obtained through this *Super-ensemble* model has enhanced the Dice-score and HD metrics for all the lesions (shown in bold font). Fig. 6 illustrates the visual quantification of the performance of the *Super ensemble-pp* (with post-processing). It highlights that the model faced challenges in identifying specific highly complex tumor lesions as depicted in Fig. 6(g), (i). However, the model demonstrated accurate segmentation of various other lesions. This visual assessment provides insights into the model's strengths and limitations in accurately identifying and segmenting regions and lesions.

Lastly, we compared the proposed method with the leading 2D or 3D models of BraTS2020 competition in Table 6. The proposed method showcased superior performance in segmenting respective lesions represented by the gray-colored regions. We can observe that the proposed methods have performed better than numerous 2D and 3D models shown in Table 6. This proves the method's ability to handle the complexities of the lesions and provide more reliable and accurate results.

4. Conclusion, limitations and future work

Here, we proposed an ensemble approach that provides robust segmentation, requires less computation, has a simple architecture, and can be broadly applied in clinical practice. The basic architecture was built on 2D U-Net. The study suggests attention mechanisms undoubtedly increase models' discriminating capability among complex lesions.

Table 5
Experimental evaluation of top performing models on BraTS2020 training set. Bold text signifies best results. Results from models were indicated using the same color for training and validation sets (shown in Table 2).

Dataset	Model Type	Model Name	Parameter Numbers in millions (M)	Dice-score			HD		
				ET	WT	TC	ET	WT	TC
Training	Triplanar Ensemble	SCSA	30.999 M	0.727	0.885	0.826	39.745	9.099	9.734
Training	Triplanar Ensemble[a]	CCSAv1[a]	31.767 M	0.727	0.892	0.829	34.538	6.828	6.128
Training	Super-Ensemble	CA-CCSA-SCSA	92.470 M	0.712	0.897	0.837	40.310	6.378	7.114
Training	Super-Ensemble[a]	CA-CCSAv1-SCSA (without postprocessing)[a]	93.500 M	0.724	0.898	0.843	39.489	5.884	5.710
Training	Super-Ensemble-pp[a]	CA-CCSAv1-SCSA (with post-processing)[a]	93.500 M	**0.736**	**0.896**	**0.841**	**35.842**	6.005	7.041

[a] Signifies a modified version of models/

Table 6
On the BraTS2020 validation dataset, the proposed method is compared with the leading models of the BraTS2020 competition. The first row displays the proposed method's specifics. The proposed method outperforms other models depicted by cells shaded in yellow. The best results are shown in boldface. The proposed method outperforms other leading models depicted by cells shaded in light yellow.([43–50])

Model name	Model Type	Dice-score			HD			Mean sensitivity ET, WT, TC	Mean specificity ET, WT, TC
		ET	WT	TC	ET	WT	TC		
Proposed method	2D Ensemble U-Net	0.713	0.873	0.778	31.563	6.347	13.274	**0.796**	0.999
Messaoudi et al. [24]	3D-U-Net	0.654	0.841	0.680	NA	NA	NA	NA	NA
Ma et al. [43]	2D-U-Net	0.704	0.879	0.773	NA	NA	NA	NA	NA
Ali et al. [26]	2D, 3D Ensemble U-Net	0.748	0.871	0.748	**3.929**	**9.428**	10.090	0.780	0.996
Agravat et al. [12]	3D-U-Net	**0.763**	0.873	0.753	27.704	07.038	10.873	0.809	0.999
Xu et al. [44]	2D-U-Net	0.673	0.861	0.704	40.608	7.942	15.750	NA	NA
Soltaninejad et al. [45]	3D-U-Net	0.660	0.870	**0.800**	47.330	6.910	7.800	0.773	**1**
Colman et al. [46]	2D-U-Net	0.676	0.886	0.672	47.620	12.110	15.740	NA	NA
Bommineni et al. [47]	3D-U-Net	0.718	0.884	0.788	30.767	4.834	9.258	NA	NA
Ballestar et al. [48]	3D-U-Net	0.720	0.840	0.790	37.970	10.930	12.240	NA	NA
Tarasiewicz et al. [49]	2D Ensemble U-Net	0.703	**0.888**	0.749	40.132	4.552	10.678	0.791	0.999
Wacker et al. [50]	2D U-Net	0.685	0.878	0.753	40.385	7.936	17.675	NA	NA
McHugh et al. [23]	2D U-Net	0.712	0.881	0.789	40.600	6.720	10.200	NA	NA

We observed sequential and concurrent combination of channel and spatial attention allows models to learn "what to see" and "where to see" from the given input images. In the paper, *CCSAv1* has performed better than *CA* and *SCSA*, and without a significant rise in several parameters. Additionally, post-processing enhances the segmentation results. Furthermore, we observed that coronal planar models could not generalize well compared to other orthogonal planar models due to the anisotropy effect. In summary, an optimized triplanar model can be combined to predict accurate segmentation, which can help experts make more informed decisions and provide timely and accurate diagnoses, even in resource-constrained environments.

The model has certain limitations. Firstly, the proposed model struggles to remove some false positive tumor legions. Second, the triplanar model intuitively fails to fully exploit the depth of information, limiting its discriminative ability. Lastly, the number of trainable parameters in the proposed model is excessive for real-time brain tumor segmentation on edge devices.

There is a significant potential for the improvement and development of the proposed methodology in the future. This includes the supplementary post-processing techniques, such as substituting ET lesions with NC based on specific threshold values identified during inference and combining segmentation with overall survival prediction [51]. Moreover, exploration of the combination of various 3D, 2D, and 2.5D U-Net holds potential for enhancing the segmentation. Additionally, integrating modalities, such as Diffusion-weighted imaging (DWI), Magnetic Resonance Spectroscopic Imaging (MRSI), and perfusion MRI, can offer supplementary biological insights. Augmenting the structural information from standard MRI scans with biological data can improve the accuracy of brain tumor segmentation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

M. Roy acknowledges the seed grant No. $ORSP/R\&D/PDPU/2019/MR-/RO051$of PDEU, India (for the computing facility), the core research grant No. $CRG/2020/000869$ of the Science and Engineering Research Board (SERB), Govt. of India and the project grant no $GUJCOST/STI/2021-22/3873$ of Gujarat Council of Science and Technology (GUJCOST), Govt. of Gujarat, India. M. S. Raval acknowledges the grant number. $GUJCOST/STI/2021-22/3858$ of Gujarat Council of Science and Technology (GUJCOST), Govt. of Gujarat, India for computing facility to support the work.

References

[1] F.B. Mesfin, M.A. Al-Dhahir, Gliomas - National Library of Medicine, 2023, https://www.nlm.nih.gov/medlineplus/glioma.html. (Accessed: 23 October 2023).

[2] S. Rajput, R. Agravat, M. Roy, M.S. Raval, Glioblastoma multiforme patient survival prediction, in: Proceedings of 2021 International Conference on Medical Imaging and Computer-Aided Diagnosis (MICAD 2021) Medical Imaging and Computer-Aided Diagnosis, Springer, 2022, pp. 47–58, http://dx.doi.org/10.1007/978-981-16-3880-0_6.

[3] P.K. Mall, P.K. Singh, S. Srivastav, V. Narayan, M. Paprzycki, T. Jaworska, M. Ganzha, A comprehensive review of deep neural networks for medical image processing: Recent developments and future opportunities, Healthc. Anal. (2023) 100216, http://dx.doi.org/10.1016/j.health.2023.100216.

[4] S. Rajput, R.A. Kapdi, M.S. Raval, M. Roy, Interpretable machine learning model to predict survival days of malignant brain tumor patients, Machine Learning: Sci. Tech. (ISSN: 2632-2153) 4 (2) (2023) 025025, http://dx.doi.org/10.1088/2632-2153/acd5a9.

[5] L. Pana, S. Moldovanu, L. Moraru, Statistical filters for processing and reconstruction of 3D brain MRI, in: 2021 25th International Conference on System Theory, Control and Computing, ICSTCC, IEEE, 2021, pp. 655–658, http://dx.doi.org/10.1109/ICSTCC52150.2021.9607186.

[6] F. Ullah, S.U. Ansari, M. Hanif, M.A. Ayari, M.E.H. Chowdhury, A.A. Khandakar, M.S. Khan, Brain MR image enhancement for tumor segmentation using 3D U-net, Sensors 21 (22) (2021) 7528, http://dx.doi.org/10.3390/s21227528.

[7] F. Isensee, P.F. Jaeger, S.A. Kohl, J. Petersen, K.H. Maier-Hein, nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation, Nature Methods 18 (2) (2021) 203–211, http://dx.doi.org/10.1038/s41592-020-01008-z.

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

[9] M. Aggarwal, A.K. Tiwari, M.P. Sarathi, A. Bijalwan, An early detection and segmentation of brain tumor using deep neural network, BMC Med. Inf. Decis. Mak. 23 (1) (2023) 1–12, http://dx.doi.org/10.1186/s12911-023-02174-8.

[10] M.M.R. Siddiquee, A. Myronenko, Redundancy reduction in semantic segmentation of 3D brain tumor MRIS, 2021, http://dx.doi.org/10.48550/arXiv.2111.00742, arXiv preprint arXiv:2111.00742.

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

[12] R.R. Agravat, M.S. Raval, 3D semantic segmentation of brain tumor for overall survival prediction, in: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part II 6, Springer, 2021, pp. 215–227, http://dx.doi.org/10.1007/978-3-030-72087-2_19.

[13] M.M. Islam, P. Barua, M. Rahman, T. Ahammed, L. Akter, J. Uddin, Transfer learning architectures with fine-tuning for brain tumor classification using magnetic resonance imaging, Healthc. Anal. (2023) 100270, http://dx.doi.org/10.1016/j.health.2023.100270.

[14] R. Ranjbarzadeh, A. Bagherian Kasgari, S. Jafarzadeh Ghoushchi, S. Anari, M. Naseri, M. Bendechache, Brain tumor segmentation based on deep learning and an attention mechanism using MRI multi-modalities brain images, Sci. Rep. 11 (1) (2021) 10930, http://dx.doi.org/10.1038/s41598-021-90428-8.

[15] Y. Yuan, Automatic brain tumor segmentation with scale attention network, in: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part I 6, Springer, 2021, pp. 285–294, http://dx.doi.org/10.1007/978-3-030-72084-1_26.

[16] M. Chi, H. An, X. Jin, K. Wen, Z. Nie, SCAR U-Net: A 3D spatial-channel attention ResU-Net for brain tumor segmentation, in: Proceedings of the 3rd International Symposium on Artificial Intelligence for Medicine Sciences, 2022, pp. 497–501, http://dx.doi.org/10.1145/3570773.3570826.

[17] X. Liu, S. Hou, S. Liu, W. Ding, Y. Zhang, Attention-based multimodal glioma segmentation with multi-attention layers for small-intensity dissimilarity, J. King Saud Univ.-Comput. Inf. Sci. 35 (4) (2023) 183–195, http://dx.doi.org/10.1016/j.jksuci.2023.03.011.

[18] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, in: Advances in Neural Information Processing Systems, vol. 30, 2017, http://dx.doi.org/10.48550/arXiv.1706.03762.

[19] W. Wang, C. Chen, M. Ding, H. Yu, S. Zha, J. Li, Transbts: Multimodal brain tumor segmentation using transformer, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24, Springer, 2021, pp. 109–119, http://dx.doi.org/10.1007/978-3-030-87193-2_11.

[20] A. Hatamizadeh, V. Nath, Y. Tang, D. Yang, H.R. Roth, D. Xu, Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images, in: International MICCAI Brainlesion Workshop, Springer, 2021, pp. 272–284, http://dx.doi.org/10.1016/j.jksuci.2023.03.011.

[21] T.H. Nguyen, C.H. Le, D.V. Sang, T. Yao, W. Li, Z. Wang, Efficient brain tumor segmentation with dilated multi-fiber network and weighted bi-directional feature pyramid network, in: 2020 Digital Image Computing: Techniques and Applications, DICTA, IEEE, 2020, pp. 1–7, http://dx.doi.org/10.1109/DICTA51227.2020.9363380.

[22] Z. Luo, Z. Jia, Z. Yuan, J. Peng, HDC-Net: Hierarchical decoupled convolution network for brain tumor segmentation, IEEE J. Biomed. Health Inf. 25 (3) (2020) 737–745, http://dx.doi.org/10.1109/JBHI.2020.2998146.

[23] H. McHugh, G.M. Talou, A. Wang, 2d Dense-UNet: A clinically valid approach to automated glioma segmentation, in: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part II 6, Springer, 2021, pp. 69–80, http://dx.doi.org/10.1007/978-3-030-72087-2_7.

[24] H. Messaoudi, A. Belaid, M.L. Allaoui, A. Zetout, M.S. Allili, S. Tliba, D. Ben Salem, P.-H. Conze, Efficient embedding network for 3D brain tumor segmentation, in: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part I 6, Springer, 2021, pp. 252–262, http://dx.doi.org/10.1007/978-3-030-72084-1_23.

[25] V. Sundaresan, G. Zamboni, P.M. Rothwell, M. Jenkinson, L. Griffanti, Triplanar ensemble U-Net model for white matter hyperintensities segmentation on MR images, Med. Image Anal. 73 (2021) 102184, http://dx.doi.org/10.1016/j.media.2021.102184.

[26] M.J. Ali, M.T. Akram, H. Saleem, B. Raza, A.R. Shahid, Glioma segmentation using ensemble of 2D/3D U-Nets and survival prediction using multiple features fusion, in: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part II 6, Springer, 2021, pp. 189–199, http://dx.doi.org/10.1007/978-3-030-72087-2_17.

[27] V. Sundaresan, L. Griffanti, M. Jenkinson, Brain tumour segmentation using a triplanar ensemble of U-Nets on MR images, in: International MICCAI Brainlesion Workshop, Springer, 2020, pp. 340–353, http://dx.doi.org/10.1007/978-3-030-72084-1_31.

[28] R. McKinley, M. Rebsamen, R. Meier, R. Wiest, Triplanar ensemble of 3D-to-2D CNNs with label-uncertainty for brain tumor segmentation, in: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 5th International Workshop, BrainLes 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 17, 2019, Revised Selected Papers, Part I 5, Springer, 2020, pp. 379–387, http://dx.doi.org/10.1007/978-3-030-46640-4_36.

[29] Y. Xie, B. Yang, Q. Guan, J. Zhang, Q. Wu, Y. Xia, Attention mechanisms in medical image segmentation: A survey, 2023, http://dx.doi.org/10.48550/arXiv.2305.17937, arXiv preprint arXiv:2305.17937.

[30] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141, http://dx.doi.org/10.48550/arXiv.1709.01507.

[31] A.G. Roy, N. Navab, C. Wachinger, Concurrent spatial and channel 'squeeze & excitation'in fully convolutional networks, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part I, Springer, 2018, pp. 421–429, http://dx.doi.org/10.1007/978-3-030-00928-1_48.

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

[33] J. Li, J. Zheng, M. Ding, H. Yu, Multi-branch sharing network for real-time 3D brain tumor segmentation, J. Real-Time Image Process. (2021) 1–11, http://dx.doi.org/10.1007/s11554-020-01049-9.

[34] C. Chen, X. Liu, M. Ding, J. Zheng, J. Li, 3D dilated multi-fiber network for real-time brain tumor segmentation in MRI, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part III 22, Springer, 2019, pp. 184–192, http://dx.doi.org/10.1007/978-3-030-32248-9_21.

[35] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J.S. Kirby, J.B. Freymann, K. Farahani, C. Davatzikos, Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features, Sci. Data 4 (1) (2017) 1–13, http://dx.doi.org/10.1038/sdata.2017.117.

[36] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, R.T. Shinohara, C. Berger, S.M. Ha, M. Rozycki, et al., Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge, 2018, http://dx.doi.org/10.48550/arXiv.1811.02629, arXiv preprint arXiv:1811.02629.

[37] B.H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, et al., The multimodal brain tumor image segmentation benchmark (BRATS), IEEE Trans. Med. Imaging 34 (10) (2014) 1993–2024, http://dx.doi.org/10.1109/TMI.2014.2377694.

[38] N.J. Tustison, B.B. Avants, P.A. Cook, Y. Zheng, A. Egan, P.A. Yushkevich, J.C. Gee, N4ITK: Improved N3 bias correction, IEEE Trans. Med. Imaging 29 (6) (2010) 1310–1320, http://dx.doi.org/10.1109/TMI.2010.2046908.

[39] K. He, X. Zhang, S. Ren, J. Sun, Delving deep into rectifiers: Surpassing human-level performance on imagenet classification, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1026–1034, http://dx.doi.org/10.1109/ICCV.2015.123.

[40] S. Rajput, R. Kapdi, M. Raval, M. Roy, Multi-view brain tumor segmentation (MVBTS): An ensemble of planar and triplanar attention UNets, Turk. J. Electr. Eng. Comput. Sci. 31 (6) (2023) 908–927, http://dx.doi.org/10.55730/1300-0632.4026.

[41] C.H. Sudre, W. Li, T. Vercauteren, S. Ourselin, M. Jorge Cardoso, Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations, in: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, QuÉBec City, QC, Canada, September 14, Proceedings 3, Springer, 2017, pp. 240–248, http://dx.doi.org/10.1007/978-3-319-67558-9_28.

[42] Z. Wu, J. Wei, J. Wang, R. Li, Slice imputation: Multiple intermediate slices interpolation for anisotropic 3D medical image segmentation, Comput. Biol. Med. 147 (2022) 105667, http://dx.doi.org/10.1016/j.compbiomed.2022.105667.

[43] S. Ma, Z. Zhang, J. Ding, X. Li, J. Tang, F. Guo, A deep supervision CNN network for brain tumor segmentation, in: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part II 6, Springer, 2021, pp. 158–167, http://dx.doi.org/10.1007/978-3-030-72087-2_14.

[44] J.H. Xu, W.P.K. Teng, X.J. Wang, A. Nürnberger, A deep supervised U-attention net for pixel-wise brain tumor segmentation, in: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part II 6, Springer, 2021, pp. 278–289, http://dx.doi.org/10.1007/978-3-030-72087-2_24.

[45] M. Soltaninejad, T. Pridmore, M. Pound, Efficient MRI brain tumor segmentation using multi-resolution encoder-decoder networks, in: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part II 6, Springer, 2021, pp. 30–39, http://dx.doi.org/10.1007/978-3-030-72087-2_3.

[46] J. Colman, L. Zhang, W. Duan, X. Ye, DR-unet104 for multimodal MRI brain tumor segmentation, in: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part II 6, Springer, 2021, pp. 410–419, http://dx.doi.org/10.1007/978-3-030-72087-2_36.

[47] V.L. Bommineni, Piecenet: A redundant unet ensemble, in: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part II 6, Springer, 2021, pp. 331–341, http://dx.doi.org/10.1007/978-3-030-72087-2_29.

[48] L.M. Ballestar, V. Vilaplana, MRI brain tumor segmentation and uncertainty estimation using 3D-unet architectures, in: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part I 6, Springer, 2021, pp. 376–390, http://dx.doi.org/10.1007/978-3-030-72084-1_34.

[49] T. Tarasiewicz, M. Kawulok, J. Nalepa, Lightweight u-nets for brain tumor segmentation, in: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part II 6, Springer, 2021, pp. 3–14, http://dx.doi.org/10.1007/978-3-030-72087-2_1.

[50] J. Wacker, M. Ladeira, J.E.V. Nascimento, Transfer learning for brain tumor segmentation, 2019, http://dx.doi.org/10.48550/arXiv.1912.12452, arXiv preprint arXiv:1912.12452.

[51] S. Rajput, M.S. Raval, A review on end-to-end methods for brain tumor segmentation and overall survival prediction, 2020, http://dx.doi.org/10.48550/arXiv.2006.01632, arXiv preprint arXiv:arXiv:2006.01632.