Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 22 No. 5 (2026) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v22i05.59425

PAPER

R-AttNet: Residual Encoder-Induced Attention Network for Brain Lesion Localization

Pradyumna Kumar Sahoo[1], **Bhramara Bar Biswal**[1](✉), **Deepak Kumar Sahoo**[2]

[1]Gandhi Institute of Engineering and Technology University, Gunupur, India

[2]Sri Sri University, Cuttack, India

bhramarabarbiswal@giet.edu

ABSTRACT

Early brain tumor localization is a crucial task for better treatment planning. Computer vision plays a significant role in vision-assisted diagnosis of tumors. In the last few decades, various vision-assisted techniques have been developed by different researchers. However, these existing techniques are incapable of accurately separating tumors of varying sizes from different modalities. Therefore, in this paper, we introduce R-AttNet, a deep learning-based residual system for accurate brain tumor localization. The developed framework has three phases of novelties: the developed residual encoder network (RAN) comprises various residual blocks that make the model computationally efficient and can extract the required multi-scale details effectively. The designed spatial attention mechanism acts as a bridge between the encoder and decoder that highlights the prominent details, which are highly essential for tumor localization. The proposed decoder network provides the extracted tumor mask while retaining the spatial dependency among the pixels efficiently. The effectiveness of the developed algorithm is corroborated using visual demonstration as well as objective analysis. The findings of this study, compared against various existing learning-based systems, may be suitable for clinical settings.

KEYWORDS

brain tumor, magnetic resonance imaging (MRI), residual encoder, attention map

1 INTRODUCTION

The detection and localization of brain lesions from magnetic resonance imaging (MRI) images are crucial for diagnosing and treating various neurological disorders. Identifying lesions early can significantly improve patient outcomes by enabling timely interventions and personalized treatment plans [1]. MRI imaging provides detailed, non-invasive views of brain structures, making it an ideal tool for detecting subtle abnormalities. However, manually analyzing these complex images is time-consuming and prone to human error. Automated detection and localization systems [2] can increase the efficiency and accuracy (ACC) of

Sahoo, P. K., Biswal, B. B., Sahoo, D. K. (2026). R-AttNet: Residual Encoder-Induced Attention Network for Brain Lesion Localization. *International Journal of Online and Biomedical Engineering (iJOE)*, 22(5), pp. 70–86. https://doi.org/10.3991/ijoe.v22i05.59425

Article submitted 2025-10-31. Revision uploaded 2026-01-28. Final acceptance 2026-01-28.

radiological assessments, which may reduce diagnostic delays and help with clinical decision-making. Additionally, accurately locating lesions is vital for surgical planning, monitoring disease progression, and assessing treatment effectiveness. As the number of neurological disorders worldwide keeps growing, developing advanced, AI-driven tools for brain lesion analysis becomes more important for supporting healthcare professionals and enhancing patient care [3].

Several AI-driven techniques have been developed by various researchers across the globe in the last decade [4], [5], [6], [7], [8], [9], [10], [11]. At the advent of the convolutional neural network (CNN), this arena has revolutionized lesion detection. Both pretrained and customized CNNs have been employed for the task. But pretrained CNNs suffer from high computational complexity and ineffective domain adaptation. So, custom CNNs have an advantage in the domain with less computational complexity. However, the existing custom CNNs suffer from poor performance due to reasons including lesion heterogeneity within the MRI scans, high computational complexity, and false positives. Therefore, in this paper, we have developed a custom CNN named R-AttNet with an encoder-decoder architecture that induces an attention mechanism for effective extraction of brain lesions from brain MRI images. The major contributions of the designed R-AttNet architecture include:

- The designed residual encoder network (RAN) uses residual blocks to extract more complex and subtle features from the MRI scans. The designed RAN network helps avoid the vanishing gradient problem, which is crucial for accurately characterizing tumors.
- The developed spatial attention network (SAN), which connects the encoder and decoder networks, enables the model to focus and highlight areas relevant to tumors.
- The proposed customized decoder network learns to emphasize features from tumor-relevant areas based on the spatial attention network. It also suppresses irrelevant information from healthy tissue, ensuring that the decoder receives a highly focused and refined signal for more precise segmentation.
- The consistent use of the GeLU (Gaussian error linear unit) activation function improves training stability and model performance, allowing for a more nuanced flow of information through the network's layers.

The rest of the paper is organized as follows: Section 2 discusses the relevant literature in the field of brain lesion extraction. The different datasets used for experimentation and validation are discussed in Section 3. Section 4 provides a detailed elaboration of the designed R-AttNet network. The subjective and objective evaluations, along with ablation experiments of R-AttNet on various datasets, are reported in Section 5. Section 6 justifies the designed R-AttNet architecture via various ablation experiments. Finally, the paper concludes with future recommendations in Section 6.

2 LITERATURE REVIEW

Automatic detection and localization of brain lesions in medical images, such as MRI, is an important job in neuro-oncology. It assists healthcare professionals in making quick and accurate diagnoses and planning treatment. In this arena, computer vision-assisted techniques play a crucial role in isolating the affected tumor

tissues from the healthy ones. Over the past decade, numerous approaches have been developed by various researchers to segregate the tumor from MRI scans. El-Shafai et al. [12] developed a hybrid localization approach combining clustering with particle swarm optimization. However, the usage of metaheuristics on full-scale images increases the time complexity of the approach and is validated on limited data. To get rid of the aforementioned limitation, Kasar et al. [13] introduced a SEGNET and UNet-based framework utilizing deep learning frameworks for the said task. However, the usage of a pretrained network limited the segmentation performance of the framework, with an overlapping DSC score of 76%, as the pretrained network does not cope with the domain shift from natural images to medical images without drastic parameter tuning. Moreover, another major concern is the number of trainable parameters, which restricts it from being deployed towards real-time applications.

To alleviate this, Jiang et al. [14] combined a custom CNN with the Swin transformer for localizing the tumors. But the experimentation is limited to only BraTS data, which indicate that the model captures the glioma tumors whose locations are easily captured within the white matter regions of the MRI scans, whereas the meningioma and pituitary tumors are generally located in the critical structures, which are not effectively located as the intensity variations are more abrupt as compared to the distinct glioma tumors in white matter. Moreover, it struggles to localize the tumors when their presence is near the edges or boundaries of the MRI scans. Later, Liu et al. [15] introduced CSWin-UNet, which is a hybrid CNN-transformer-based U-shaped architecture integrated self-attention mechanism for tumor localization from MRI scans. However, the integration of the transformer within the framework increases the parameter to 23.57 M, limiting its utility towards real-time edge device deployment. Islam et al. [16] introduced CoST-UNet, a customized CNN-transformer framework utilizing the Swin transformer, integrating the U-Net for tumor extraction. The approach utilizes preprocessing with significant parameters; 31 M compared to its baseline Swin-UNet. Also, the model struggles with poor-contrast clinical images, as it uses adaptive histogram equalization as one of the preprocessing steps before analyzing the region of interest. It is observed from these aforementioned discussions that the CNNs capture the local context, whereas the transformers capture the global features, which create semantic disparities for the network to effectively capture these features, which would contribute towards tumor localization. To get rid of this, TransSea [17] provided a hybrid CNN-transformer combination that captures semantically aware features from the MRI scans. But limited validation with the glioma tumor raises a concern about its applications to other tumor categories. Another approach, also utilizing a similar context, TransResUNet [18], also demonstrated tumor extraction capability for glioma tumors, but over 250 epochs of training with limited data make it appear overfitted. Further, ETUNet [19] utilized a token learner with spatial attention and a channel attention module for glioma tumor extraction. However, it suffers from limited generalization with increased trainable parameters. Ghazouani et al. [20] utilized the Swin transformer with a local self-attention mechanism for glioma tumors but provided a DSC value below 90%. Most approaches developed so far using a hybrid CNN customized configuration suffer from high trainable parameters, making the model more complex, and are primarily developed for glioma tumors, which may not apply to diverse datasets.

MM-LinkNet [21], developed for gliomas, attained a DSC of 77% with significant parameters and time complexity. Later, combining both supervised DNN with clustering, the approach [4] obtained an average DSC value of 84.66%,

leading to ineffective tumor localization. Angona et al. [22] introduced a Swin Transformer-based attention mechanism that improved the DSC value to 87.23%. To get rid of bulkier architecture, Hernandez-Gutierrez et al. [23] developed a lightweight U-Net with 2M parameters but attained a DSC value of 86.0%. It also fails to capture the tumor regions when tumors are comparatively small. Therefore, based on the above literature, it can be inferred that there is a need for a customized CNN that can accommodate tumors of varying sizes with reduced computational complexity and improved performance metrics, which must be validated across diverse and challenging medical imaging datasets. To address this, we have introduced R-AttNet, which is a light-weight attention-based CNN framework that can effectively localize tumors for diverse MRI datasets. In an attention U-Net, unlike the U-Net, the skip connection between the encoder and decoder networks is established through an attention gate instead of a direct skip connection. Further, the attention U-Net does not use the residual encoder like the U-Net. Furthermore, in U-Net or Attention U-Net, ReLU activations are used. On the other hand, the proposed R-AttNet uses non-linear GeLU instead of ReLU activation so that it can prevent the dead neurons and handle non-linearity smoothly to segment the fine details. Also, the proposed model uses residual connections in the encoder network for effective deep feature learning, strong gradient flow, and high training stability. Furthermore, the SAN used in the proposed framework increases the convergence speed with a lower false positive rate for tumor detection.

3 DATASET DESCRIPTION

To validate the performance of the proposed network, three benchmark datasets named Figshare [24], BraTS 2020 [25], and BraTS 2021 [26] are considered in this paper. The Figshare Dataset comprises 3,064 T1-weighted contrast-enhanced (T1-CE) MRI slices, representing three primary tumor types, including meningioma, glioma, and pituitary tumors, across 233 patients. Each image slice is a 2D axial slice from clinical MRI scans, intended for research in the detection, classification, and segmentation of brain tumors. The dataset exhibits significant diversity in tumor size, shape, intensity, and location, reflecting real-world clinical variations. All images are preprocessed and standardized to ensure consistency, with annotations generally provided as binary masks distinguishing tumor from background. Due to its manageable size and high-quality annotations, this dataset is widely used for developing and benchmarking deep learning algorithms in brain tumor segmentation. The BraTS 2020 dataset includes 369 pre-operative multimodal MRI scans (T1, T1Gd, T2, and T2-FLAIR) with expert-annotated sub-regions: necrotic/non-enhancing core, edema, and enhancing tumor, enabling segmentation and survival prediction tasks. The BraTS 2021 dataset expands to nearly 2,000 cases with the same MRI modalities and introduces a radio-genomic task for MGMT promoter methylation prediction. Both BraTS datasets are pre-processed and provided in NIfTI format for benchmarking robust tumor analysis.

4 PROPOSED METHOD

In this work, we have introduced an artificial learning system named ResAttNet comprising various stages for MRI image segmentation. In this work, the developed

RAN is capable of extracting significant multi-scale details from the MRI image. The designed SAN can maintain the correlation between the encoder and decoder networks, which is suitable for retaining masks with reduced noise. The proposed decoder framework can upsample the in-depth details while preserving the spatial coherence among the pixels. The detailed description of the developed algorithm is depicted in Figure 1.

4.1 RAN

Residual encoder network is the first stage of the proposed ResAttNet model, primarily responsible for extracting significant features from the input medical image while preserving spatial context. For the input MRI scan with the dimension $x \in \Re^{H \times W \times 1}$, the process begins with an initial batch normalization, followed by a GeLU activation and max pooling operation that produces output $RE_{i1} \in \Re^{\frac{H}{2} \times \frac{W}{2} \times 1}$, REi_1 with dimension. This combination ensures input normalization, introduces smooth non-linearity for better gradient flow, and decreases the spatial resolution, thereby making the network more computationally efficient.

The RAN is structured as a series of three residual blocks, each consisting of convolutional layers, batch normalization, and GeLU activations. In the residual blocks, the skip connections are used to alleviate the vanishing gradient problem, accelerate convergence, and enable deeper network training. By adding the input of a block directly to its output, the model retains low-level details while simultaneously learning high-level abstract features. In the developed system, the first residual block consists of different convolutional layers with 3×3, 64 filters, batch normalization, and GeLU activations that map from the image space dimension $\frac{H}{2} \times \frac{W}{2} \times 1$ denoted as RE_{i1} to the feature space dimension of $RE_{o1} \in \Re^{\frac{H}{2} \times \frac{W}{2} \times 64}$. As the network progresses through the second and third residual blocks, which are a similar arrangement to the first residual block, the spatial dimensions of the feature maps are reduced from $\frac{H}{2} \times \frac{W}{2} \times 64$ to $\frac{H}{8} \times \frac{W}{8} \times 256$ while the channel depth is increased from 64 to 256. The outcome of the first residual block (RE_{o1}) is processed through the GeLU activation function that produces details (RE_{i2}) given to the second residual block. For the feature maps RE_{i2}, the second residual block generates diverse detail indicated as RE_{o2}, which is further processed through the GeLU activation function that generates significant features (RE_{i3}). For the input (RE_{i3}), the third residual block produces high-level features (RE_{o3}) that are passed through the GeLU activation function. For the given MRI input scan with dimensions $H \times W \times 1$, the proposed RAN can extract multi-scale details with dimensions $\frac{H}{8} \times \frac{W}{8} \times 256$.

4.2 SAN

Spatial attention network plays a key role in refining the encoded details by selectively emphasizing informative spatial regions while suppressing less relevant background information. Unlike conventional feature extraction that treats all spatial locations equally, the SAN dynamically learns where to focus within the feature maps, which is especially important in medical image segmentation tasks

where target structures are often small, irregular, and embedded within complex backgrounds.

The designed SAN is integrated between the RAN and the decoder network, comprising four blocks where each block receives multi-scale details from the encoder and processes them through a dual-branch attention mechanism. One branch captures average-pooled spatial information, while the other extracts max-pooled spatial details, ensuring that both subtle and dominant features are considered. These feature descriptors are concatenated along the channel dimension, then passed through convolutional layers followed by sigmoid activation, which learns attention coefficients that highlight discriminative spatial regions. In the designed SAN, the first and second block's convolution layer contains 256 filters of size 3×3 with a stride = 1. However, the third and fourth block's convolution layer contains 128 and 64 filters of size 3×3 with a stride of 2, respectively. As a result, the designed SAN network can handle multi-scale contextual information. Since details from different residual blocks of the encoder framework are given into the SAN, the designed SAN network learns to attend to both global contextual information (from deeper layers) and local details (from shallower layers). This multi-level integration improves the designed model's ability to segment regions with better ACC, even when intensity contrasts are poor.

By incorporating hierarchical detail maps with spatial attention, the designed SAN framework ensures that the subsequent decoder network receives feature representations that are both semantically rich and spatially precise. This significantly enhances the final mask prediction by sharpening boundaries and reducing false positives.

4.3 Decoder network

The developed decoder network serves as the final stage of the proposed architecture, responsible for generating the segmentation mask from the refined feature representations obtained through the RAN and SAN architectures. Its primary function is to gradually recover the spatial resolution of the encoded feature maps while preserving the contextual and structural information learned during encoding and attention refinement.

The designed decoder architecture consists of five blocks, where the first four blocks sandwich Transposed convolutional (Transposed conv), convolutional, and GeLU activations. Initially, the extracted multi-scale features with sizes $\frac{H}{8} \times \frac{W}{8} \times 256$ are given to the first block SAN to produce significant details. Again, the outcome of the developed encoder model, combined with SAN's first block outcome, is fed to the first block of the decoder network that consists of Transposed conv and convolutional layers with 256 filters of size 3×3, which further refines the spatial details and reduces semantic information loss introduced during the downsampling in the encoder. Then the outcomes of the decoder's first block, combined with SAN's second block, result in feature maps (d_1) of dimensions $\frac{H}{8} \times \frac{W}{8} \times 256$ that are both semantically meaningful and spatially accurate. Such fusion is particularly beneficial in

medical image segmentation, where retaining sharp boundaries and subtle tissue details is essential.

Further, the d_1 details are fed to the second block of the decoder's network, where Transposed conv and convolutional layers having 128 filters of size 3×3. The outcomes of the decoder's second block are combined with SAN's third block output, which provides refined details denoted as d_2 of dimensions $\frac{H}{8} \times \frac{W}{8} \times 256$. Similarly, the feature maps d_2 are given to the third block of the decoder framework, where the Transposed conv with a stride of 2 and convolutional layers consisting of 64 filters with size 3×3. The results obtained by the decoder's third block, combined with SAN's fourth block outcome, generate significant feature maps indicated as d_3 of dimension $\frac{H}{4} \times \frac{W}{4} \times 64$. Furthermore, feature maps d_3 are given to the fourth block of the decoder framework, where the Transposed conv with a stride of 4 and convolutional layers consisting of 64 filters with size 3×3, which restored to the original input resolution, provides 64 distinct details (d_4) of dimension $H \times W \times 64$. At the final stage, the decoder framework applies a pixel classification head, consisting of a convolutional layer with a single filter of size 3×3 followed by a softmax function and a pixel classification layer, to generate the final segmentation mask.

The complete operation of the R-AttNet is described through a set of mathematical operations as follows: The outputs from residual encoder blocks are passed through the GeLU activation functions and are given as the input to the successive encoders by using equations (1), (2), and (3).

$$RE_{i2} = GeLU(RE_{o1}) \tag{1}$$

$$RE_{i3} = GeLU(RE_{o2}) \tag{2}$$

$$RE_{i4} = GeLU(RE_{o3}) \tag{3}$$

Where RE_{o1}, RE_{o2}, and RE_{o3} are the outputs from residual encoder modules.

Further, RE_{i2} and RE_{i3} act as the inputs to the upper two SAN modules with RE_{i4} as the input to the lower two SAN modules shown in Figure 1. The outputs from the SAN modules are given in equations (4), (5), (6), and (7).

$$d_0 = (RE_{i4} \oplus s_1) \epsilon \, \mathfrak{K}^{\frac{H}{8} \times \frac{W}{8} \times 256} \tag{4}$$

$$d_1 = (D_0 \oplus s_2) \epsilon \, \mathfrak{K}^{\frac{H}{8} \times \frac{W}{8} \times 256} \tag{5}$$

$$d_2 = (D_1 \oplus s_3) \epsilon \, \mathfrak{K}^{\frac{H}{8} \times \frac{W}{8} \times 128} \tag{6}$$

$$d_3 = (D_2 \oplus s_2) \epsilon \, \mathfrak{K}^{\frac{H}{4} \times \frac{W}{4} \times 64} \tag{7}$$

Where s_1, s_2, s_3, and s_4 are the outputs from the SAN modules, with the outputs from the lower three decoder modules being D_0, D_1, and D_2.

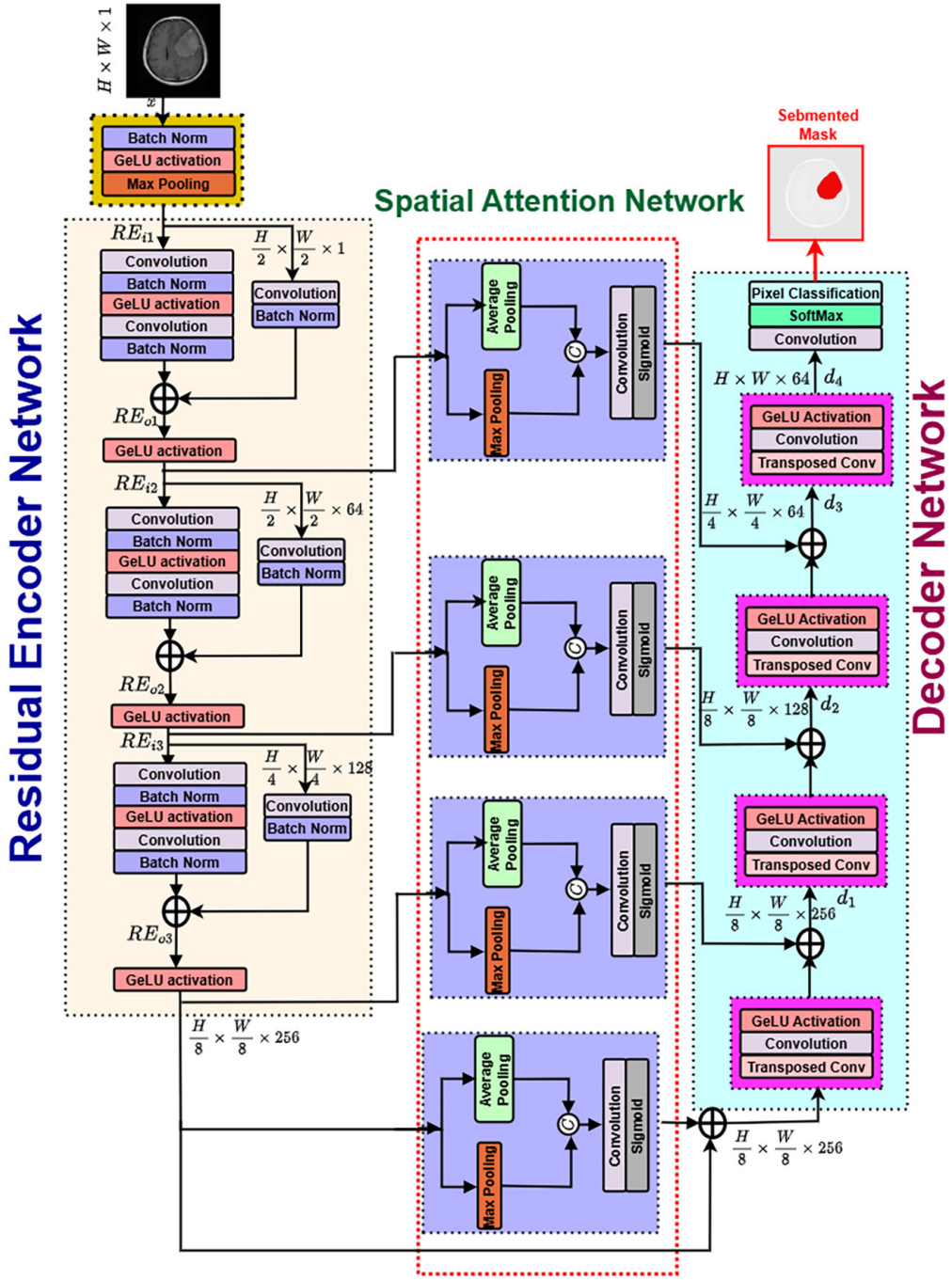

Fig. 1. Architecture of the proposed method

5 RESULTS AND DISCUSSIONS

The proposed R-AttNet is designed to automatically extract the brain tumor lesion using 2D-MRI scans. The model is trained and tested in the MATLAB environment using a Core i5 processor and an 8 GB GPU. For the training, the learning rate is chosen as 0.001 with 60 epochs and a mini-batch size of 8. To validate the

performance of the present model, particularly, the Figshare brain tumor data are tested in three separate stages of experiment with varying training and testing splits. During the first phase, 90% of the images were trained, and the remaining 10% were tested. In the second and third training-testing stages, 80%/20% and 70%/30% training-testing splits were used, respectively. At each stage of the image level, the dataset was split and randomly performed such that the training and testing sets were mutually exclusive. The data splitting strategy is used the same way throughout all experimental sessions, and the randomization of the data is also regulated to prevent data leakage. This multi-stage evaluation protocol is chosen to determine the strength and the generalization power of the proposed model when different levels of training data are used. The training accuracies found with 90%, 80%, and 70% training images are 99.2%, 98.92%, and 98.86%, respectively, which are very close to each other. Hence, it can be observed that the model performs properly with variations in the number of training samples. For testing purposes, we have chosen the model with 90% training images. Once the model is trained, it is further tested on all the images in the dataset.

5.1 Performance parameters

To validate the effectiveness of the developed R-AttNet in localizing brain tumor lesions, four performance indices [5] are used in the paper: ACC, Dice similarity coefficient (DSC), Jaccard index (JI), and sensitivity (SN). In the present paper, performance is evaluated using multiple metrics because each captures a different aspect of ACC. Pixel ACC reflects the overall proportion of correctly classified pixels, while SN highlights how well the model detects true lesion pixels, ensuring that critical regions are not missed. The JI measures the overlap between predicted and ground truth regions, providing a strict assessment of segmentation quality. Similarly, the Dice coefficient evaluates spatial similarity, balancing false positives and false negatives for robust comparison. Together, these metrics give a comprehensive understanding of both pixel-level correctness and region-level overlap in segmentation performance.

5.2 Visual demonstration

The visual demonstration of the network performance is illustrated in Figures 2 and 3. In Figure 2, nine image samples are considered as part of the visual demonstration. Similarly, in Figure 3, four image samples are taken. The first three samples in Figures 2 and 3 represent the scans for meningioma tumor lesions, the next three samples represent Glioma tumor lesions, and the final three scans are for Pituitary tumor lesions. Further, Figures 2a and 3a represent the original images, Figures 2b and 3b represent the annotated ground truths, Figures 2c and 3c give the semantic segmented outputs from the network, Figures 2d and 3d depict the segmented binary masks, and finally, Figures 2e and 3e give the masked images. From Figures 2c, 3c, 2d, 3d, 2e, and 3e, it can be observed that the proposed R-AttNet is highly efficient in localizing the brain tumor lesion.

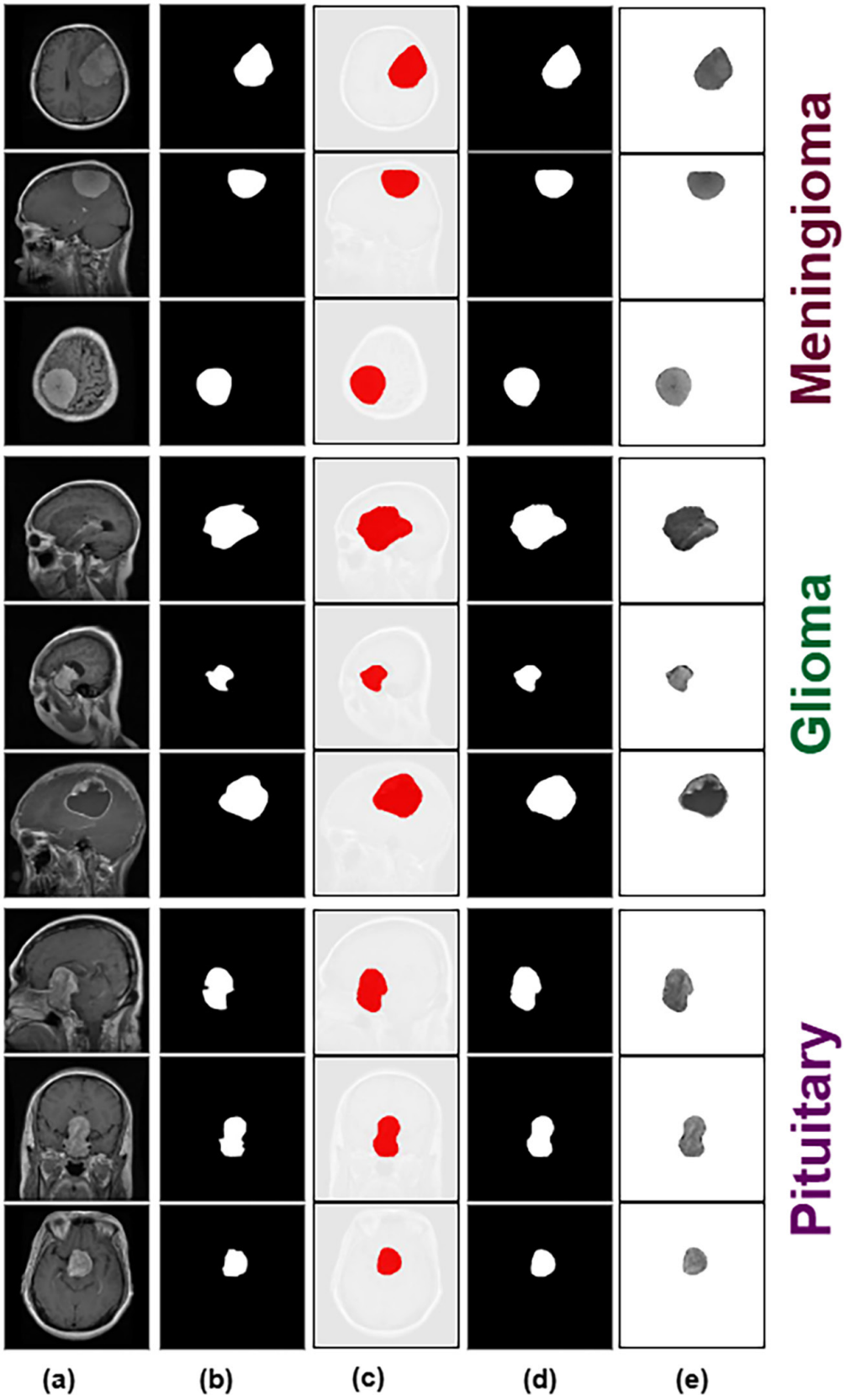

Fig. 2. Outputs of the proposed ResAttNet using the Figshare brain tumor dataset. (a) Original image, (b) Ground truth, (c) Semantic segmentation output, (d) Binary extracted mask, (e) Masked output

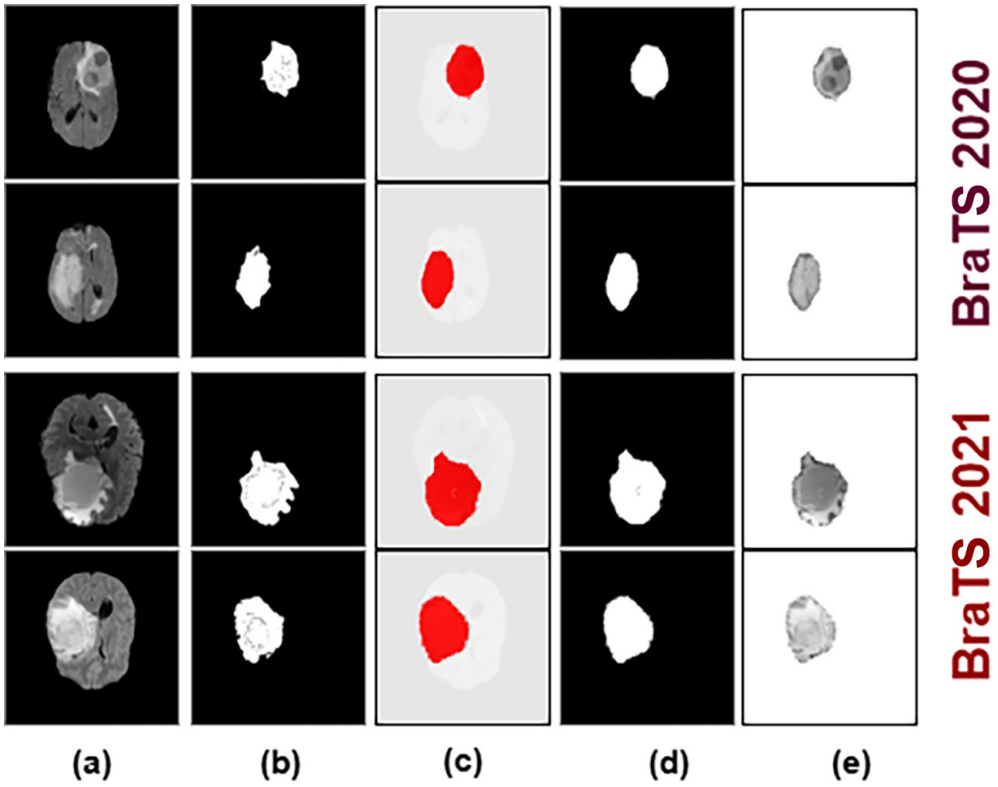

Fig. 3. Outputs of the proposed ResAttNet using BraTS datasets: (a) Original image, (b) Ground truth, (c) Semantic segmentation output, (d) Binary extracted mask, (e) Masked output

5.3 Quantitative analysis

For the quantitative evaluation of the proposed R-AttNet, the performance metrics, including ACC, DSC, JI [27], and SN, are calculated for the complete dataset and reported in the Table 1.

Table 1. Quantitative presentation of performance parameters using different benchmark datasets

Tumor Type	Datasets	ACC (%)	DSC (%)	JI (%)	SN (%)
Meningioma	Figshare	99.11	90.04	81.76	86.57
Glioma		98.93	89.96	81.91	86.56
Pituitary		99.27	89.75	81.38	86.69
Average		99.10	89.91	81.68	86.60
Glioma	BraTS 2020	98.92	89.58	81.14	88.64
Glioma	BraTS 2021	98.65	90.61	82.82	89.92

Table 1 presents the quantitative performance evaluation of the proposed R-AttNet using the Figshare and BraTS brain tumor dataset. The evaluation is carried out on three major tumor categories, namely meningioma, glioma, and pituitary tumors, with performance reported in terms of ACC, DSC, JI, and SN. From Table 1, it can be observed that the proposed network achieves consistently high ACC across all tumor types, with values ranging from 98.65% for glioma to 99.27% for pituitary tumors. The DSC values remain close to 90% for all three categories, indicating a strong overlap

between the predicted and ground-truth tumor regions. Similarly, the JI values range between 81.14% and 82.82%, further confirming the reliable segmentation performance of the model. SN values also remain balanced across the tumor types, ranging between 86.56% and 89.92%, which demonstrates the ability of the model to effectively capture lesion regions without significant under-segmentation. On average, the proposed R-AttNet achieves an overall ACC of 99.1%, DSC of 89.91%, JI of 81.68%, and SN of 86.60% for the Figshare dataset. Similarly, the average ACC, DSC, JI, and SN using the BraTS 2020 dataset are 98.92%, 89.58%, 81.14%, and 88.64%, respectively. Further, the same using the BraTS 2021 dataset are 98.65%, 90.61%, 82.82%, and 89.92%, respectively. These results highlight the robustness and generalization capability of R-AttNet, ensuring consistent performance across different tumor categories with diverse brain tumor datasets. Such findings emphasize that the integration of residual encoders with attention mechanisms improves the discriminative power of the network, enabling precise and reliable brain lesion localization.

Further, Table 2 provides a comprehensive comparison of the proposed R-AttNet with several recent state-of-the-art brain tumor lesion extraction models using both Figshare and BraTS datasets. In this work, the results reported for various SOTA methods are collected from the published manuscript, which indicates that we have compared the results obtained by the proposed techniques against different existing methods without changing the environmental settings. The evaluation is based on three key performance indicators: ACC, DSC, and SN. Early approaches such as PSO + IFCM (2022) achieved relatively low overlap scores with a DSC of 35.18% and a SN of only 37.92%, indicating limitations of traditional optimization-based clustering methods.

Table 2. Comprehensive comparison of the proposed R-AttNet with several recent state-of-the-art brain tumor lesion extraction models using the Figshare and BraTS datasets

Methods	Datasets	Year	ACC (%)	DSC (%)	SN (%)
PSO + IFFCM [12]	Figshare	2022	93.47	35.18	37.92
UNET and SEGNET [13]		2022	97.80	76.00	–
SwinBTS [14]		2023	–	88.91	80.50
CSWin-UNet [15]		2024	–	86.49	79.90
CoST-UNet [16]		2024	–	87.63	81.30
TransSea [17]		2024	–	88.21	81.55
TransResUNet [18]		2024	–	89.25	82.10
ETUNet [19]		2024	–	85.49	77.80
Swin Transformer [20]		2024	–	88.32	81.85
R-AttNet (Ours)			**99.1**	**89.91**	**86.60**
Encoder-Based Link-Net [21]	BraTS 2020	2023	–	77.33	76.62
DNN + Clustering [4]		2023	–	84.88	92.75
Residual U-Net + Attention + Swin Transformer [22]		2025	–	87.23	–
R-AttNet (Ours)			**98.92**	**89.58**	**88.64**
Light Weight U-Net [23]	BraTS 2021	2024	–	86.00	85.60
MVSI-Net [28]		2024	–	81.9	81.1
2D Ensemble U-Net [29]		2024	–	87.30	79.6
R-AttNet (Ours)			**98.65**	**90.61**	**89.92**

Similarly, UNET and SEGNET (2022) reported improvements with DSC reaching 76% and ACC up to 97.8%, reflecting the efficiency of encoder–decoder architectures. More advanced models from 2023 onwards, such as SwinBTS, achieved a DSC of 88.91% and an ACC of 88.2%, showing the potential of transformer-based backbones for medical image segmentation. In 2024, several high-performing networks were introduced, including CSWin-UNet (DSC 86.49%), CoST-UNet (DSC 87.63%), and TransSea (DSC 88.21%), which improved both overlap and SN metrics through hybrid convolution–transformer designs. TransResUNet and ETUNet further boosted DSC values to 89.25% and 88.49%, respectively, with SN surpassing 82% in some cases, highlighting the benefits of residual learning and attention mechanisms. The Swin Transformer (2024) also maintained a balanced performance with DSC 88.32% and SN 81.85%, reinforcing the growing role of transformer-based architectures. Compared to these methods, the proposed R-AttNet achieves a superior balance across all metrics, with an ACC of 99.1%, DSC of 89.91%, and SN of 86.60%, outperforming most existing CNN and transformer-based models using the Figshare dataset.

Furthermore, while comparing with SOTA approaches using the BraTS 2020 dataset, the proposed R-AttNet achieves a DSC of 89.58% and SN of 88.64%, which are better than the Encoder Based Link-Net, Residual U-Net + Attention + Swin Transformer model, and DNN + Clustering approach. Similarly, using the BraTS 2021 dataset, the calculated DSC and SN for the proposed model are 90.61% and 89.92%, which are superior to Light Weight U-Net, MVSI-Net, and 2D Ensemble U-Net that are shown in Table 2. Further, Table 3 represents a comparison of the space complexity of the proposed R-AttNet with other methods. From Table 3, it can be seen that the space complexity for the developed model is low compared to the other existing SOTA methods. However, the space complexity of the designed network is comparable to the MVSI-Net [28] method, whereas the performance of MVSI-Net [28] is low as compared to the proposed network.

Table 3. Comparison of the space complexity of the proposed R-AttNet with recent state-of-the-art methods

Methods	Year	Space Requirement
UNET and SEGNET [13]	2022	28.25 M
CSWin-UNet [15]	2024	23.57 M
CoST-UNet [16]	2024	22.99 M
Residual U-Net + Attention + Swin Transformer [22]	2025	6.1 M
MVSI-Net [28]	2024	4.76 M
2D Ensemble U-Net [29]	2024	31 M
R-AttNet (Ours)		**5.1 M**

Importantly, R-AttNet attains the highest ACC among all compared methods while maintaining competitive DSC and SN, reflecting its robustness in tumor boundary delineation. These results demonstrate that combining residual encoders with attention modules enables more discriminative feature extraction, ensuring precise and reliable lesion localization. Furthermore, the consistent improvement across metrics suggests that R-AttNet is not only accurate but also clinically practical, as it reduces false negatives while maintaining high segmentation fidelity. Overall, the comparative study establishes R-AttNet as a strong advancement in brain tumor segmentation, offering superior ACC and competitive overlap performance relative to recent literature.

5.4 Ablation study

For the effectiveness of the selection of various components in the proposed R-AttNet, an ablation study is performed. In the first ablation analysis is conducted based on the selection of the activation function, whereas in the second study, it is conducted based on the usage and effect of the attention network on the performance of the proposed architecture.

Table 4. Ablation study on the selection of components in the proposed network using the Figshare dataset

Network Type	ACC (%)	DSC (%)
Network with ReLU activation	96.28	84.21
Network without SAN	95.76	81.82
Network with SAN and GeLU activation (R-AttNet)	**99.1**	**89.91**

The ablation study presented in Table 4 demonstrates the effectiveness of different components in the proposed R-AttNet architecture. When ReLU activation is used instead of GeLU, the model achieves an ACC of 96.28% and a DSC of 84.21%, indicating that while ReLU provides reasonable performance, it is not optimal for this task. Similarly, when the SAN is removed, the performance further declines to 95.76% ACC and 81.82% DSC, highlighting the crucial role of SAN in capturing spatial dependencies and enhancing segmentation quality. In contrast, the complete proposed R-AttNet, which integrates SAN with GeLU activation, achieves the highest performance with 99.1% ACC and 89.91% DSC. This significant improvement clearly validates the combined contribution of SAN and GeLU activation in boosting both classification ACC and segmentation consistency, establishing the superiority of the proposed architecture over its ablated variants.

6 CONCLUSION

The proposed R-AttNet demonstrates strong and consistent performance for brain tumor lesion extraction across key evaluation metrics. Compared with recent CNN- and transformer-based architectures from 2022 to 2024, R-AttNet achieves the highest ACC of 99.1%, while also delivering a competitive Dice score of 89.91% and SN of 86.60%. These results highlight its ability to balance precise tumor boundary delineation with reliable lesion detection. Unlike traditional clustering and early encoder–decoder models, R-AttNet leverages residual encoders and attention modules to enhance feature representation and capture fine-grained spatial details. The model not only surpasses several advanced architectures such as CSWin-UNet, TransSea, and Swin Transformer but also remains lightweight and computationally efficient. Its robustness across tumor types makes it suitable for clinical use. Furthermore, the model reduces false negatives, which is crucial in medical imaging tasks where missed lesions can have severe implications. Overall, R-AttNet establishes itself as a state-of-the-art solution in brain tumor localization. However, the developed R-AttNet architecture needs to be tested on an unseen setup. Further, the effectiveness of the developed algorithm is yet to be validated in real-time clinical settings.

7 REFERENCES

[1] F. J. Dorfner, J. B. Patel, J. Kalpathy-Cramer, E. R. Gerstner, and C. P. Bridge, "A review of deep learning for brain tumor analysis in MRI," *npj Precis. Oncol.*, vol. 9, no. 1, p. 2, 2025. https://doi.org/10.1038/s41698-024-00789-2

[2] A. K. Sahoo, P. Parida, M. K. Panda, K. Muralibabu, and A. S. Mohanty, "MultiTumor Analyzer (MTA-20–55): A network for efficient classification of detected brain tumors from MRI images," *Biocybern. Biomed. Eng.*, vol. 44, no. 3, pp. 617–634, 2024. https://doi.org/10.1016/j.bbe.2024.06.003

[3] M. Ariful Islam, M. F. Mridha, M. Safran, S. Alfarhood, and M. Mohsin Kabir, "Revolutionizing brain tumor detection using explainable AI in MRI images," *NMR Biomed.*, vol. 38, no. 3, p. e70001, 2025. https://doi.org/10.1002/nbm.70001

[4] A. K. Sahoo, P. Parida, and K. Muralibabu, "Hybrid deep neural network with clustering algorithms for effective gliomas segmentation," *Int. J. Syst. Assur. Eng. Manag.*, vol. 15, no. 3, pp. 964–980, 2024. https://doi.org/10.1007/s13198-023-02183-w

[5] A. K. Sahoo, P. Parida, K. Muralibabu, and S. Dash, "An improved DNN with FFCM method for multimodal brain tumor segmentation," *Intell. Syst. with Appl.*, vol. 18, p. 200245, 2023. https://doi.org/10.1016/j.iswa.2023.200245

[6] A. K. Sahoo and P. Parida, "Automatic clustering based approach for brain tumor extraction," *Journal of Physics: Conference Series*, vol. 1921, no. 1, p. 012007, 2021. https://doi.org/10.1088/1742-6596/1921/1/012007

[7] A. K. Sahoo and P. Parida, "A clustering based approach for meningioma tumors extraction from brain MRI images," in *2020 IEEE International Symposium on Sustainable Energy, Signal Processing and Cyber Security (iSSSC)*, IEEE, 2020, pp. 1–5. https://doi.org/10.1109/iSSSC50941.2020.9358849

[8] A. S. Mohanty, K. C. Patra, and P. Parida, "Toddler ASD classification using machine learning techniques," *Int. J. Online Biomed. Eng.*, vol. 17, no. 7, pp. 156–171, 2021. https://doi.org/10.3991/ijoe.v17i07.23497

[9] R. Ray, S. Jena, P. Parida, L. Dash, and S. K. Biswal, "Empowering diabetic eye disease detection: Leveraging differential evolution for optimized convolution neural networks," *Int. J. Online Biomed. Eng.*, vol. 20, no. 10, pp. 86–100, 2024. https://doi.org/10.3991/ijoe.v20i10.49187

[10] T. A. Ton Komar Azaharan, A. K. Mahamad, S. Saon, Muladi, and S. W. Mudjanarko, "Investigation of VGG-16, ResNet-50 and AlexNet performance for brain tumor detection," *Int. J. Online Biomed. Eng.*, vol. 19, no. 8, pp. 97–109, 2023. https://doi.org/10.3991/ijoe.v19i08.38619

[11] C. B. Chandrakala, S. Pooja, C. Pujari, S. Ketavarapu, S. Awatramani, and S. Gohil, "Health-Lens: A health diagnosis companion," *Int. J. Interact. Mob. Technol.*, vol. 19, no. 12, pp. 68–102, 2025. https://doi.org/10.3991/ijim.v19i12.51525

[12] W. El-Shafai *et al.*, "Hybrid segmentation approach for different medical image modalities," *Comput. Mater. Contin.*, vol. 73, no. 2, pp. 3455–3472, 2022. https://doi.org/10.32604/cmc.2022.028722

[13] P. E. Kasar, S. M. Jadhav, and V. Kansal, "MRI modality-based brain tumor segmentation using deep neural networks," *Research Square*, 2021. https://doi.org/10.21203/rs.3.rs-496162/v1

[14] Y. Jiang, Y. Zhang, X. Lin, J. Dong, T. Cheng, and J. Liang, "SwinBTS: A method for 3D multimodal brain tumor segmentation using swin transformer," *Brain Sci.*, vol. 12, no. 6, p. 797, 2022. https://doi.org/10.3390/brainsci12060797

[15] X. Liu, P. Gao, T. Yu, F. Wang, and R.-Y. Yuan, "CSWin-UNet: Transformer UNet with cross-shaped windows for medical image segmentation," *Inf. Fusion*, vol. 113, p. 102634, 2025. https://doi.org/10.1016/j.inffus.2024.102634

[16] M. R. Islam, M. Qaraqe, and E. Serpedin, "CoST-UNet: Convolution and swin transformer based deep learning architecture for cardiac segmentation," *Biomed. Signal Process. Control*, vol. 96, p. 106633, 2024. https://doi.org/10.1016/j.bspc.2024.106633

[17] Y. Liu, Y. Ma, Z. Zhu, J. Cheng, and X. Chen, "TransSea: Hybrid CNN–Transformer with semantic awareness for 3-D brain tumor segmentation," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 16–31, 2024. https://doi.org/10.1109/TIM.2024.3413130

[18] N. Rasool, J. Iqbal Bhat, N. Ahmad 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

[19] W. Zhang, S. Chen, Y. Ma, Y. Liu, and X. Cao, "ETUNet: Exploring efficient transformer enhanced UNet for 3D brain tumor segmentation," *Comput. Biol. Med.*, vol. 171, p. 108005, 2024. https://doi.org/10.1016/j.compbiomed.2024.108005

[20] F. Ghazouani, P. Vera, and S. Ruan, "Efficient brain tumor segmentation using Swin transformer and enhanced local self-attention," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 19, no. 2, pp. 273–281, 2023. https://doi.org/10.1007/s11548-023-03024-8

[21] G. Ramasamy, T. Singh, and X. Yuan, "Multi-modal semantic segmentation model using encoder based link-net architecture for BraTS 2020 challenge," *Procedia Comput. Sci.*, vol. 218, pp. 732–740, 2023. https://doi.org/10.1016/j.procs.2023.01.053

[22] T. M. Angona and M. R. H. Mondal, "An attention based residual U-Net with swin transformer for brain MRI segmentation," *Array*, vol. 25, p. 100376, 2025. https://doi.org/10.1016/j.array.2025.100376

[23] F. D. Hernandez-Gutierrez *et al.*, "Brain tumor segmentation from optimal MRI slices using a lightweight U-Net," *Technologies*, vol. 12, no. 10, p. 183, 2024. https://doi.org/10.3390/technologies12100183

[24] J. Cheng, "Brain tumor dataset," Figshare, 2024. [Online]. Available: https://figshare.com/articles/dataset/brain_tumor_dataset/1512427

[25] 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*, 2019. https://doi.org/10.48550/arXiv.1811.02629

[26] Q.-H. Trinh, T.-H. N. Mau, R. Zosimov, and M.-V. Nguyen, "EfficientNet for Brain-Lesion classification," *arXiv preprint arXiv:2208.04616*, 2022. https://doi.org/10.48550/arXiv.2208.04616

[27] A. K. Sahoo, P. Parida, and K. Muralibabu, "Effective use of clustering techniques for brain tumor segmentation," in *2023 IEEE 3rd International Conference on Applied Electromagnetics, Signal Processing, & Communication (AESPC)*, IEEE, 2023, pp. 1–5. https://doi.org/10.1109/AESPC59761.2023.10390467

[28] J. Sun *et al.*, "MVSI-Net: Multi-view attention and multi-scale feature interaction for brain tumor segmentation," *Biomed. Signal Process. Control*, vol. 95, p. 106484, 2024. https://doi.org/10.1016/j.bspc.2024.106484

[29] S. Rajput, R. Kapdi, M. Roy, and M. S. Raval, "A triplanar ensemble model for brain tumor segmentation with volumetric multiparametric magnetic resonance images," *Healthc. Anal.*, vol. 5, p. 100307, 2024. https://doi.org/10.1016/j.health.2024.100307

8 AUTHORS

Pradyumna Kumar Sahoo is a Ph.D. scholar in the Department of Computer Science and Engineering at the Gandhi Institute of Engineering and Technology University, Gunupur, Odisha. His research focuses on developing machine learning algorithms for large-scale data analytics, with a particular emphasis on graph neural

networks and their applications in social network analysis. He holds a Master's degree in Computer Science from the Biju Patnaik University of Technology (BPUT), Odisha. Pradyumna has showcased his research at prestigious conferences such as NeurIPS and ICML. In addition to his academic pursuits, he mentors undergraduate students and actively contributes to initiatives that promote women's participation in technology (E-mail: pradyumna.sahoo@giet.edu).

Bhramara Bar Biswal is currently serving as an Associate Professor in the School of Engineering and Technology, GIET University. He possesses a distinguished academic background in Computer Science and Engineering, with over 23 years of teaching and research experience. He holds a Ph.D. and an M.Tech. in Computer Science from Berhampur University. Dr. Biswal has published more than 20 research papers in reputed SCOPUS, SCIE, and UGC CARE-listed journals, as well as in IEEE conferences, Taylor & Francis publications, and book chapters. He also holds patents to his credit. At present, seven Ph.D. scholars are pursuing their research under his supervision and are nearing the completion of their theses. He has been associated with IIT Bombay and IIT Kharagpur as an NMEICT coordinator. His research interests include Artificial Intelligence and Machine Learning (AI/ML), Deep Learning, and Big Data (MapReduce). He is an editor of an IEEE journal and a Life Member of several professional bodies, including CSI and ISTE. Throughout his career, Dr. Biswal has served in various academic and administrative capacities, such as Assistant Professor, Associate Professor, and Head of Department in different engineering institutions. He has also contributed as a member of research and founding committees, as well as NAAC and NBA committees at the university level (E-mail: bhramarabarbiswal@giet.edu).

Deepak Kumar Sahoo earned his Master in computer science and engineering from the International Institute of Information Technology (IIIT-Bhubaneswar). He earned his Ph.D. in computer science and engineering from the International Institute of Information Technology (IIIT-Bhubaneswar). He has worked for Odia-vertical at IIIT-Bhubaneswar in a Consortia project headed by IIIT-Bombay having the project title "Cross Lingual Information Access." He has over 15 years of teaching and research experience in various organizations. He has authored 34 publications, out of which 19 are papers in different journals indexed in SCI & Scopus, 11 are conference and 4 book chapters (E-mail: deepak.s@srisriuniversity.edu.in).