JAISCR, 2027, Vol. 17, No. 1, pp. 85 – 129
10.2478/jaiscr-2026-0025

SOFAN-UNET: A SHORT OPTIMIZATION APPROACH FOR ATTENTION NETWORKS IN U-NET USING PARTICLE SWARM STRATEGY FOR PRECISE BRAIN TUMOR SEGMENTATION

Shoffan Saifullah[1,2,*], Rafał Dreżewski[1], Anton Yudhana[3], Wahyu Caesarendra[4,5], Enny Itje Sela[6]

[1]*Faculty of Computer Science, AGH University of Krakow, al. Adama Mickiewicza 30, 30-059 Krakow, Poland*

[2]*Department of Informatics, Universitas Pembangunan Nasional Veteran Yogyakarta, Jl. Babarsari 2 Yogyakarta, 55281 Yogyakarta, Indonesia*

[3]*Department of Electrical Engineering, Universitas Ahmad Dahlan, Jl. Pramuka No.5F, Pandeyan, Umbulharjo, 55161 Yogyakarta, Indonesia*

[4]*Faculty of Engineering Technology and Science, Higher Colleges of Technology, Ras Al Khaimah, Sultan Al Kabeer East Street, Seih Al Araibi, P.O. Box 4793, United Arab Emirates*

[5]*Department of Electrical Automation Engineering, Faculty of Vocation, Institut Teknologi Sepuluh Nopember, Jl. Raya ITS, Surabaya 60111, Indonesia*

[6]*Department of Information Technology, Universitas Teknologi Yogyakarta, Jl. Siliwangi (Ringroad Utara), Jombor, 55285 Yogyakarta, Indonesia*

[*]*E-mail: saifulla@agh.edu.pl*

Submitted: 12th October 2025; Accepted: 11th May 2026

Abstract

Accurate segmentation of brain tumors in magnetic resonance imaging (MRI) is essential for clinical diagnosis and treatment planning. However, variations in tumor morphology, location, and modality-specific appearance make the task highly complex for conventional deep learning approaches. This study introduces SOFAN-UNet, a novel segmentation framework that improves tumor localization and boundary delineation through hybrid attention mechanisms and bio-inspired optimization. The proposed SOFAN-UNet architecture builds on the traditional U-Net structure and incorporates a hybrid attention mechanism combining Squeeze-and-Excitation (SE) and Self-Attention modules. A short optimization strategy, implemented via Particle Swarm Optimization (PSO), is used to dynamically calibrate attention gate parameters, including scale factors and filter sizes, within a constrained search space. This enables efficient learning of discriminative tumor features while preserving spatial detail through optimized skip connections. The model was evaluated across three benchmark datasets—FBTS, BraTS 2021, and BraTS

2018—using DSC, JI, HD, ASSD, AUC, and MCC metrics. Ablation studies and five-fold cross-validation were conducted to assess robustness. SOFAN-UNet achieved state-of-the-art performance, with Dice scores reaching 0.9685 and Jaccard Index up to 0.9389. It demonstrated strong generalization across MRI modalities and tumor types, maintained low boundary error ($HD < 2.5$), and achieved AUC values above 0.999. Ablation results confirmed the effectiveness of PSO-based calibration and attention integration, especially for heterogeneous tumor regions. The SOFAN-UNet model offers a highly accurate and computationally efficient framework for brain tumor segmentation. Its robustness across datasets and imaging modalities supports its potential for clinical deployment and further extension to subregion tumor analysis.

Keywords: brain tumor segmentation, u-net, attention mechanism, particle swarm optimization, deep learning

1 Introduction

Brain tumors remain among the most critical neurological conditions, requiring a timely and accurate diagnosis to improve treatment outcomes [1, 2]. Magnetic Resonance Imaging (MRI) is the clinical standard for brain tumor visualization due to its superior soft tissue contrast and non-invasive nature [3, 4]. Accurate segmentation of tumor regions in MRI is crucial for a range of clinical applications, including surgical planning, radiation therapy, and treatment monitoring [5, 6]. However, manual segmentation is time-consuming, prone to inter-observer variability, and difficult to scale, necessitating the development of automated and reliable computational approaches [7, 8, 9].

Convolutional neural networks (CNNs), especially U-Net and its variants [10], have become the backbone of medical image segmentation due to their encoder-decoder structure and effective preservation of spatial information through skip connections [11, 12, 13]. Despite their success, conventional U-Net architectures typically rely on fixed attention modules or predefined skip pathways, which may not generalize well across diverse tumor morphologies [14, 15, 16]. Fixed attention gates are often inadequate for capturing the complex spatial and textural variations observed in multi-modal MRI scans, leading to decreased performance in challenging segmentation cases [17, 18, 19, 20].

To overcome these limitations, we propose SOFAN-UNet—a novel U-Net-based architecture augmented with a short optimization strategy for attention learning. "Short optimization" refers to a lightweight yet powerful Particle Swarm Optimization (PSO) process that dynamically calibrates attention gate parameters—such as scale factors and filter sizes—within a constrained search space. This approach balances optimization effectiveness with computational efficiency [21], allowing the model to adaptively emphasize relevant tumor features across imaging conditions without a large amount of overhead.

SOFAN-UNet also integrates a hybrid attention mechanism that fuses Squeeze-and-Excitation (SE) modules [22, 23] for channel-wise feature recalibration with Self-Attention blocks that capture long-range spatial dependencies [24, 25]. These modules are strategically distributed throughout both the encoder and decoder, enhancing the network's ability to localize tumors with high precision. Furthermore, optimized skip connections facilitate effective fusion of hierarchical features while preserving anatomical details critical for accurate boundary delineation.

The key contributions of this study are:

1. We propose **SOFAN-UNet**, a novel segmentation model that combines PSO-driven short optimization with a hybrid attention mechanism to adaptively tune attention gates and improve spatial feature learning.
2. We design an attention structure that combines SE and Self-Attention modules to jointly model inter-channel relevance and spatial context, enhancing tumor boundary recognition.
3. We implement a constrained PSO framework to efficiently tune attention gate parameters, reduc-

ing parameter redundancy, and improving segmentation performance.

4. We validate SOFAN-UNet across three benchmark datasets—FBTS, BraTS 2021, and BraTS 2018—demonstrating superior results in Dice Similarity Coefficient (DSC), Jaccard Index (JI), AUC, and boundary-based metrics compared to state-of-the-art methods.

The remainder of this paper is organized as follows: Section 2 reviews the relevant literature on brain tumor segmentation and attention-based deep learning models. Section 3 details the SOFAN-UNet architecture, including the attention modules and PSO-based optimization. Section 4 outlines the experimental setup and presents quantitative and qualitative results across datasets. Section 5 discusses the findings, ablation experiments, and limitations. Finally, Section 6 concludes the paper and suggests directions for future work.

2 Related Works

Recent advances in brain tumor segmentation have been largely driven by convolutional neural networks (CNNs), with U-Net and its derivatives forming the architectural backbone of most state-of-the-art methods [26, 27, 28]. The encoder-decoder structure of U-Net, with symmetric skip connections, effectively captures contextual and spatial information—making it highly suitable for medical imaging tasks [11, 29, 12, 30]. Nevertheless, conventional U-Net configurations often employ fixed attention mechanisms or uninformed skip pathways [31, 11, 32], which limit their ability to focus on the salient features of the tumor, especially in cases of variable morphology, poor contrast, or complex anatomical surroundings [33, 34, 35].

To address these challenges, attention-based CNNs have been proposed. Attention U-Net [18, 36, 37, 38, 39] introduced trainable soft attention gates within skip connections, enabling selective information flow and suppressing background noise. Further refinements, such as UNet++ [40, 41, 42] and Attention ResUNet [43, 44], enhance connectivity by using nested dense paths or attention-coupled residual blocks. Although these models improve localization performance, most rely on static or heuristically chosen attention parameters, which may not optimally generalize to the diverse spatial profiles of brain tumors.

Hybrid attention approaches offer promising directions. SE-UNet [22, 45, 46, 47] leverages Squeeze-and-Excitation blocks to recalibrate inter-channel dependencies, improving contrast discrimination. More recently, transformer-based models like TransUNet [48, 49, 13, 50, 51] and Swin-UNet [52, 53, 54] integrate self-attention to model global spatial relationships. Although effective, these transformer-based architectures often require substantial computational resources and large annotated datasets, posing barriers to clinical adoption.

In addition to architectural innovations, optimization strategies have been introduced to improve CNN performance [55]. Evolutionary algorithms [56, 57, 58, 59, 60, 61] such as Particle Swarm Optimization (PSO) and Genetic Algorithm (GA) have been applied to tune learning rates, dropout rates, or filter sizes. For example, PSO-UNet [21] demonstrated that PSO can improve U-Net convergence and segmentation accuracy by optimizing global hyperparameters. However, such approaches primarily target high-level configurations and do not fine-tune internal attention modules during training, which limits adaptability.

A noticeable gap remains in dynamically calibrating attention mechanisms through optimization. Fixed attention configurations may underperform on heterogeneous or subtle tumor appearances [42, 62]. In contrast, our approach introduces a constrained PSO-based short optimization that tunes the scale factors and the dimensions of the kernel within the attention gates during training. This method balances adaptivity and computational cost by limiting the search space, offering dynamic module-level refinement without burdening the training pipeline.

Furthermore, our proposed SOFAN-UNet employs a hybrid attention system—fusing Squeeze-and-Excitation and Self-Attention blocks—across encoder and decoder layers. This dual-mode attention enables both local channel enhancement and long-range spatial context modeling, which is particularly beneficial for segmenting diffuse tumor boundaries seen in LGG or irregular Glioma cases.

In summary, while existing works have explored either advanced attention designs or global

optimization independently, few have unified both in a coherent framework. SOFAN-UNet addresses this gap by incorporating hybrid attention with PSO-driven short optimization, offering a scalable and adaptable solution for brain tumor segmentation. Its performance across diverse datasets, including FBTS, BraTS 2021, and BraTS 2018, substantiates its effectiveness and potential for clinical integration.

3 Method

This section introduces the proposed SOFAN-UNet architecture, which combines a U-Net backbone with a hybrid attention mechanism and a constrained Particle Swarm Optimization (PSO)-based short optimization strategy. The method is designed to enhance both the learning precision and adaptability of segmentation in complex brain tumor scenarios. The key innovations include a dual-attention mechanism (SE and Self-Attention) and a dynamic PSO-based adjustment of attention gate parameters within a limited search space. The architecture is evaluated using three diverse datasets: FBTS, BraTS 2021, and BraTS 2018.

3.1 Dataset and Preprocessing

This study utilizes three benchmark datasets for brain tumor segmentation: the Figshare Brain Tumor Segmentation (FBTS) dataset [63] and the BraTS (Brain Tumor Segmentation) 2021 [64] and 2018 [65, 66, 67] datasets. These datasets cover various tumor types, imaging modalities, and segmentation challenges, enabling a comprehensive evaluation of the proposed SOFAN-UNet model.

Datasets Overview

- **FBTS Dataset:** This dataset comprises T1-weighted contrast-enhanced MRI slices of three tumor types: Meningioma, Glioma, and Pituitary. Each image is accompanied by a binary segmentation mask delineating the tumor region. The dataset contains single-channel images with varying contrast, tumor location, and size, making it ideal for evaluating single-class segmentation performance.
- **BraTS Datasets (2021 and 2018):** These datasets consist of multi-modal MRI scans, including native T1, post-contrast T1 (T1CE), T2, and FLAIR sequences. Each volume contains voxel-wise annotations for different tumor subregions (enhancing tumor, edema, and necrosis). For this study, we consider the whole-tumor segmentation, combining the subregions into a unified binary mask. The datasets include both high-grade glioma (HGG) and low-grade glioma (LGG) cases, allowing the model's performance to be assessed across tumor severity levels and structural complexity.

Image Normalization and Resizing

All MRI slices and corresponding masks were preprocessed to ensure consistency across datasets [68, 69]:

- **Normalization:** Each image $I(x,y)$ was normalized to the range $[0,1]$ by dividing each pixel by 255 (Eq. 1).

$$I_{\text{norm}}(x,y) = \frac{I(x,y)}{255} \tag{1}$$

- **Resizing:** To conform to the network input size, all images and masks were resized to 256×256 pixels. Bilinear interpolation was used for the images, whereas nearest-neighbor interpolation was applied to the masks to preserve the integrity of the label.
- **Channel Expansion:** Single-channel grayscale images were expanded into three channels by replication, creating a pseudo-RGB input suitable for convolutional feature extraction.

Data Augmentation Strategy

To enhance the robustness and generalization capability of the model, we implemented a multi-transform data augmentation pipeline [70, 71, 72]. Each transformation T is applied probabilistically and aims to simulate real-world variability in spatial orientation, illumination, and morphology. The applied transformations include:

1. **Horizontal and Vertical Flipping:** Introduced to account for anatomical symmetry. The flipped image I' is defined as Eq. 2.

$$\begin{aligned} I'_{\text{flip}}(x,y) &= I(W-x-1,y) \\ &\text{or} \quad I(x,H-y-1) \end{aligned} \tag{2}$$

where W and H are the width and height of the image.

2. **Rotation:** Random rotations within $\pm 45^\circ$ simulate different patient orientations. Given a rotation angle θ, the transformation matrix is given by Eq. 3.

$$T_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{3}$$

3. **Brightness and Contrast Adjustment:** Adjusts the image I using scaling coefficients α and bias β (Eq. 4).

$$I'_{\text{contrast}} = \alpha I + \beta, \quad \text{with } \alpha \in [0.8, 1.2],\ \beta \in [-0.2, 0.2] \tag{4}$$

4. **Gaussian Blurring:** Applied with a variable kernel size $k \in [3,7]$, simulating motion blur and noise. The blurred image I' is given by:

$$I' = I * G_\sigma, \quad \text{where } G_\sigma \text{ is the Gaussian kernel} \tag{5}$$

5. **Elastic Transformation:** This warps the image using random displacement fields D_x and D_y smoothed by a Gaussian filter, mimicking tissue deformations (Eq. 6).

$$I' = \text{Elastic}(I, \alpha, \sigma) \tag{6}$$

6. **Grid Distortion:** Distorts the image by moving pixels on a regular grid. The deformation is computed using sinusoidal displacement of the grid nodes.

7. **Random Resized Cropping:** Simulates partial views by cropping and resizing. The crop region is sampled with scale $s \in [0.78, 1.0]$ and fixed aspect ratio $r = 1.0$ (Eq. 7).

$$I' = \text{ResizeCrop}(I, s, r) \tag{7}$$

These augmentations are applied jointly during training, and each transformation contributes to the diversity of training samples, reducing overfitting, and enhancing generalization to unseen data.

Dataset Samples

Figure 1 presents visual examples from both datasets. The top row illustrates T1-weighted images with binary masks for the three tumor classes in FBTS. The lower section includes multimodal MRI sequences (T1, T2, T1CE, FLAIR) along with ground-truth masks from the BraTS 2021 and BraTS 2018 datasets.

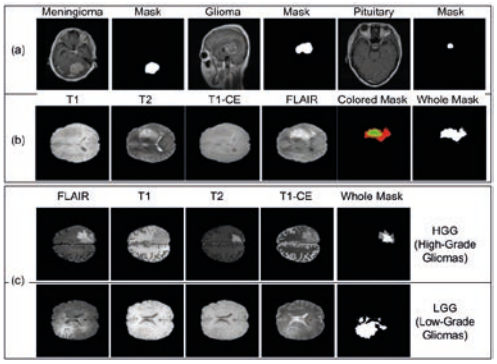


Figure 1. Representative images and segmentation masks from the utilized datasets. (a) FBTS dataset samples for Meningioma, Glioma, and Pituitary tumors, along with their binary masks. (b) BraTS 2021 dataset slices for T1, T2, T1CE, and FLAIR modalities with corresponding ground-truth tumor masks. (c) BraTS 2018 samples from HGG and LGG cases across modalities, along with their respective whole-tumor masks.

3.2 SOFAN-UNet Architecture

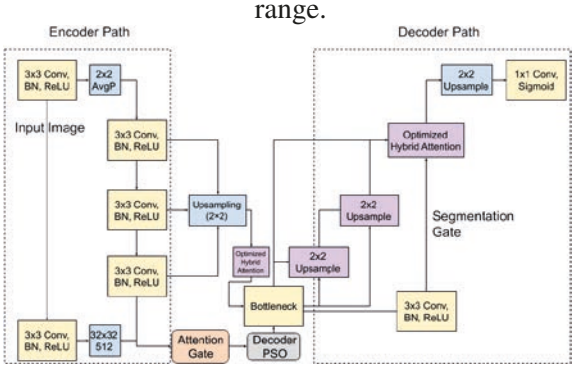
The SOFAN-UNet model is a modified U-Net architecture designed to improve brain tumor segmentation performance by incorporating hybrid attention gates tuned using a short-range Particle Swarm Optimization (PSO) strategy. This approach enables the model to emphasize essential features while suppressing irrelevant regions, particularly around complex tumor boundaries. The network is composed of three essential components: (1) an encoder–decoder backbone, (2) PSO-optimized hybrid attention gates, and (3) a fusion strategy to combine spatial and channel-wise attention information effectively.

3.2.1 Encoder and Decoder Blocks

The encoder-decoder architecture of SOFAN-UNet (Figure 2) follows a symmetric U-shaped topology, designed to extract hierarchical features

from brain MRI scans and progressively reconstruct accurate segmentation masks. This configuration is illustrated in Figure 2, where the left side represents the encoder path, and the right side mirrors it, serving as the decoder. The architectural modifications introduced in SOFAN-UNet – including PSO-optimized attention filters at skip connections – are also marked, distinguishing it from the conventional U-Net design.

(a) SOFAN-UNet architectural design. The model consists of an encoder–decoder backbone with convolutional blocks, upsampling layers, and hybrid attention modules at each skip connection. A Particle Swarm Optimization (PSO) module is used to optimize attention parameters in a short range.



(b) Workflow of SOFAN-UNet on brain MRI segmentation. This diagram shows spatial resolution transitions across the encoder–decoder structure, optimized attention gate placements, and segmentation output overlaying the tumor region.

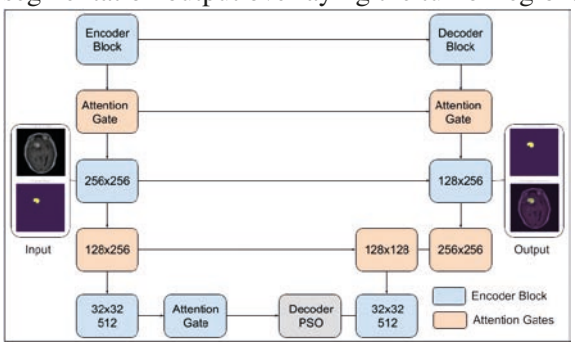


Figure 2. Overview of the proposed SOFAN-UNet framework. (a) illustrates the architectural details and module interconnections, while (b) presents the data flow and output visualization on a sample input.

In the encoder path, the network processes the input image $\mathbf{X}_0 \in \mathbb{R}^{256\times256\times3}$ through four convolutional stages, each composed of a two-dimensional convolution, batch normalization, a nonlinear activation function, and spatial downsampling. Specifically, each block applies a convolution operation with a kernel size of 3×3, followed by Batch Normalization (BN) and Rectified Linear Unit (ReLU) activation [73, 74] to form the initial transformation (Eq. 8).

$$\mathbf{F}_{\text{enc}}^{(l)} = \text{ReLU}\left(\text{BN}\left(\text{Conv}_{3\times3}\left(\mathbf{F}_{\text{enc}}^{(l-1)}\right)\right)\right) \tag{8}$$

After feature transformation, average pooling is performed using a 2×2 filter with a stride 2 to reduce the spatial resolution [75] by half (Eq. 9).

$$\mathbf{F}_{\text{pool}}^{(l)} = \text{AvgPool}_{2\times2}\left(\mathbf{F}_{\text{enc}}^{(l)}\right) \tag{9}$$

This downsampling increases the receptive field of the subsequent layers, allowing the model to capture contextual features relevant to the global tumor structure. The number of feature channels is doubled after each encoder block [76, 77], resulting in the following dimensional progression: $64 \rightarrow 128 \rightarrow 256 \rightarrow 512$, while the spatial resolution decreases from 256×256 to 32×32.

At the bottom of the network lies the bottleneck layer, which further transforms the compressed features [18, 78]. This stage consists of an additional 3×3 convolution, batch normalization, and ReLU activation to deepen semantic encoding (Eq. 10).

$$\mathbf{F}_{\text{bottleneck}} = \text{ReLU}\left(\text{BN}\left(\text{Conv}_{3\times3}\left(\mathbf{F}_{\text{pool}}^{(4)}\right)\right)\right) \tag{10}$$

The bottleneck encodes high-level abstract features that describe the tumor region as a whole, serving as a bridge between the encoder and the decoder.

The path of the encoder, shown on the right side of Figure 2, reverses the transformations made in the encoder. It restores spatial resolution using upsampling layers and integrates detailed spatial information from the encoder via skip connections. Each decoder block begins with nearest-neighbor interpolation to double the spatial resolution (Eq. 11).

$$\mathbf{F}_{\text{up}}^{(l)} = \text{Up}_{2\times2}\left(\mathbf{F}_{\text{dec}}^{(l+1)}\right) \tag{11}$$

This is followed by concatenation with the corresponding encoder feature map $\mathbf{F}_{\text{enc}}^{(l)}$, which has been filtered through an attention gate (Eq. 12).

$$\mathbf{F}_{\text{concat}}^{(l)} = \left[\mathbf{F}_{\text{up}}^{(l)} \| \mathbf{F}_{\text{enc}}^{\text{att}(l)}\right] \tag{12}$$

The concatenated feature map is processed by 3×3 convolution, batch normalization, and ReLU activation (Eq. 13).

$$\mathbf{F}_{\text{dec}}^{(l)} = \text{ReLU}\left(\text{BN}\left(\text{Conv}_{3\times3}\left(\mathbf{F}_{\text{concat}}^{(l)}\right)\right)\right) \tag{13}$$

As the decoder progresses, the number of feature channels is halved at each stage: from 512 to 256, then to 128, and finally to 64. This channel reduction ensures efficient memory usage while maintaining the necessary detail for fine-grained segmentation.

The final segmentation map is generated through a 1×1 convolutional layer that reduces the output tensor to a single channel. A sigmoid activation function is applied to map each pixel value to a probability between 0 and 1 (Eq. 14).

$$\hat{Y} = \sigma\left(\text{Conv}_{1\times1}\left(\mathbf{F}_{\text{dec}}^{(1)}\right)\right) \tag{14}$$

This probability map represents the likelihood of each pixel being part of the tumor region.

Importantly, the encoder and decoder paths are tightly coupled through attention-gated skip connections, enabling effective reuse of encoder features. These attention gates, discussed in Section 3.2.2, filter out irrelevant features before they are passed to the decoder, ensuring that only semantically and spatially relevant information is used [79]. This design addresses a major limitation in traditional U-Net architectures, where unfiltered skip connections may introduce noise or irrelevant texture patterns into the decoding process.

In general, the encoder-decoder configuration in SOFAN-UNet is carefully designed to balance the extraction of high-level tumor semantics with the retention of fine-grained anatomical structure. By applying attention before fusion and preserving both low-level and deep contextual information, the network achieves superior performance in segmenting complex tumor shapes, particularly for indistinct or diffuse tumor boundaries.

The general architectural design of the proposed SOFAN-UNet is illustrated in Figure 3.2.1. The network follows a U-shaped encoder-decoder structure, where the encoder progressively extracts hierarchical features through convolutional and downsampling operations, and the decoder reconstructs the segmentation mask using upsampling and skip connections. Unlike standard U-Net architectures, SOFAN-UNet integrates PSO-optimized hybrid attention gates into skip connections, enabling selective feature fusion [79] that emphasizes tumor-relevant regions. The bottleneck layer, situated between the encoder and the decoder, processes deep semantic features that encode the global context of the tumor.

Complementing this architectural view, Figure 3.2.1 presents a functional flow diagram of the segmentation process. This includes resolution transitions across stages, attention gate placements, and sample input and output MRI slices. The decoder receives selectively filtered features via attention gates whose internal parameters (such as scale factor and filter size) are dynamically optimized by the PSO controller. This short optimization strategy ensures efficient adaptation to data-specific characteristics, improving tumor localization and boundary sharpness. Together, these visualizations provide a comprehensive understanding of SOFAN-UNet's structural and operational design.

3.2.2 Hybrid Attention Gate

To enhance the relevance of features passed through skip connections in the U-Net structure, SOFAN-UNet incorporates a *hybrid attention gate* in each encoder–decoder connection. The purpose of this module is to selectively emphasize informative spatial and channel-wise features while suppressing irrelevant or noisy activations. This is especially important in medical image segmentation, where subtle anatomical structures may be lost without proper focus mechanisms.

The hybrid attention gate in SOFAN-UNet combines two attention paradigms: channel attention through a squeeze-and-excitation (SE) operation [80, 46, 81, 45] and spatial attention through a gating signal from the decoder [82, 83, 84]. The fusion of these two mechanisms ensures that both semantically significant channels and spatially relevant regions are emphasized during feature transfer

across the network.

Let $\mathbf{x} \in \mathbb{R}^{H\times W\times C}$ represent the feature map of the encoder, and let $\mathbf{g} \in \mathbb{R}^{H'\times W'\times C'}$ be the corresponding decoder gating signal. The channel attention mechanism first performs global average pooling on $\mathbf{x}$ to capture global channel statistics (Eq. 15).

$$z_c = \frac{1}{H\cdot W}\sum_{i=1}^{H}\sum_{j=1}^{W} x_{i,j,c}, \quad \forall c \in \{1,\ldots,C\} \tag{15}$$

This produces a channel descriptor vector $\mathbf{z} \in \mathbb{R}^C$, which is then passed through a bottleneck consisting of two fully connected layers with ReLU and sigmoid activations (Eq. 16).

$$s = \sigma(W_2 \cdot \delta(W_1 \cdot \mathbf{z})) \tag{16}$$

The output vector $s \in [0,1]^C$ serves as a set of channel-wise importance weights and is applied to the original encoder features by multiplication (Eq. 17).

$$\mathbf{x}_{\text{SE}} = \mathbf{x} \cdot s \tag{17}$$

To capture spatial attention, both $\mathbf{x}$ and the decoder signal $\mathbf{g}$ are first passed through 1×1 convolutions, batch normalization, and ReLU activation to align them in dimension and emphasize learned features (Eq. 18).

$$\begin{aligned}\tilde{\mathbf{x}} &= \text{ReLU}(\text{BN}(\text{Conv}_{1\times1}(\mathbf{x}))),\\ \tilde{\mathbf{g}} &= \text{ReLU}(\text{BN}(\text{Conv}_{1\times1}(\mathbf{g})))\end{aligned} \tag{18}$$

The spatial attention map α is then computed using a sigmoid-activated 1×1 convolution applied to the sum of transformed feature maps (Eq. 19).

$$\alpha = \sigma(\text{Conv}_{1\times1}(\tilde{\mathbf{x}} + \text{Up}(\tilde{\mathbf{g}}))) \tag{19}$$

Here, $\text{Up}(\cdot)$ denotes the spatial upsampling of $\tilde{\mathbf{g}}$ to match the resolution of $\tilde{\mathbf{x}}$, and $\alpha \in [0,1]^{H\times W}$ highlights spatially relevant locations.

The final attention-refined encoder feature is obtained by fusing both attention mechanisms (Eq. 20).

$$\mathbf{x}_{\text{att}} = \alpha \cdot \mathbf{x}_{\text{SE}} \tag{20}$$

As shown in Figure 3.2.1, hybrid attention gates are placed at each skip connection between the encoder and decoder stages. These gates selectively refine encoder features before fusion, enabling the model to suppress irrelevant activations and enhance focus on tumor regions. Their strategic placement ensures an effective transfer of both low-level spatial and high-level contextual features, which is essential for accurate tumor segmentation.

The internal structure of the proposed hybrid attention gate is illustrated in Figure 3. The figure demonstrates how channel attention is computed via a squeeze-and-excitation (SE) operation and how spatial attention is derived from the interaction between encoder features and the decoder gating signal. The final output of the attention gate is obtained by element-wise multiplication of both channel- and spatial-attended maps. This dual-attention mechanism ensures that both the *where* (spatial) and *what* (channel) attention are learned dynamically for each input. Such integration enables SOFAN-UNet to more accurately delineate complex and diffuse tumor boundaries in brain MRI images.

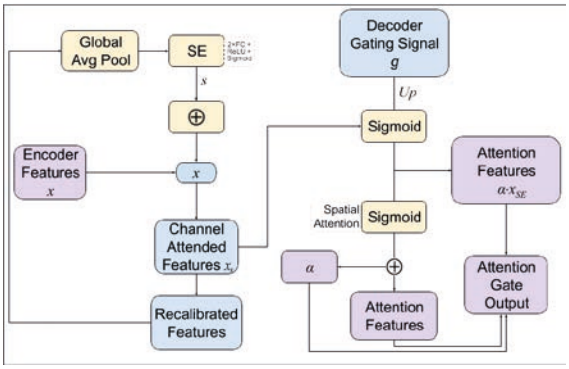


Figure 3. Internal structure of the hybrid attention gate used in SOFAN-UNet. Encoder features $\mathbf{x}$ undergo channel-wise recalibration using a squeeze-and-excitation (SE) block, while spatial attention is computed from the gating signal $\mathbf{g}$ in the decoder. The fusion of both mechanisms results in refined attention features that highlight tumor-relevant regions.

3.2.3 Particle Swarm Optimization of Attention Gates

To enhance the adaptability and performance of the hybrid attention gates in SOFAN-UNet, we introduce a short optimization strategy based on Particle Swarm Optimization (PSO). This approach allows the model to dynamically determine optimal values for internal attention parameters, such as the scale factor for feature compression and the filter size for attention gating, without requiring exhaustive manual tuning.

PSO is a population-based stochastic optimization algorithm inspired by the social behavior of bird flocking and fish schooling [85]. Each potential solution, called a particle, explores the search space by updating its velocity and position based on its own best-known position and the best-known positions of its neighbors [86, 87, 88]. The optimization proceeds iteratively, evaluating the fitness of each particle using a defined objective function [89].

Let the particle vector be defined as $\mathbf{p} = [\phi_s, \phi_f]$, where ϕ_s is the scale factor applied in the SE block, and ϕ_f is the filter size multiplier used in the attention convolutions. For each candidate solution, the attention gates are instantiated with the given parameters, and the full SOFAN-UNet model is trained for a limited number of epochs to evaluate its validation performance. The fitness function $\mathcal{F}$ is defined as the Dice similarity coefficient (DSC) on the validation set (Eq. 21).

$$\mathcal{F}(\phi_s, \phi_f) = \frac{2 \cdot |\hat{Y} \cap Y|}{|\hat{Y}| + |Y|} \tag{21}$$

where $\hat{Y}$ and Y are the predicted and ground-truth segmentation masks, respectively.

The optimization process follows the standard PSO update rules (Eq. 22 and 23).

$$\begin{aligned} v_i(t+1) = w \cdot v_i(t) + c_1 \cdot r_1 \cdot (p_i^{\text{best}} - x_i(t)) \\ + c_2 \cdot r_2 \cdot (g^{\text{best}} - x_i(t)) \end{aligned} \tag{22}$$

$$x_i(t+1) = x_i(t) + v_i(t+1) \tag{23}$$

where v_i is the velocity of the i-th particle, x_i is its current position (parameter values), p_i^{best} is the best-known position of the particle, g^{best} is the global best position, and w, c_1, and c_2 are the inertia weight and acceleration coefficients, respectively. The random numbers r_1 and r_2 are sampled uniformly from $[0, 1]$.

After several iterations, the particle with the highest validation DSC provides the optimal configuration ϕ_s^*, ϕ_f^*, which is then used to instantiate the final SOFAN-UNet model. This process is computationally efficient, since it restricts the search to a short range and uses partial training runs to estimate performance.

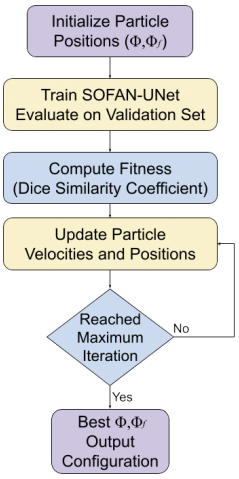


Figure 4. Flowchart of the PSO optimization process used to tune the hybrid attention gate parameters in SOFAN-UNet. The process begins with the random initialization of particle positions that represent the scale factor Φ_s and the filter size Φ_f. Each particle's fitness is evaluated using the Dice similarity coefficient on a validation set. The swarm iteratively updates particle velocities and positions until a stopping criterion is met, outputting the optimal configuration for attention gating.

As illustrated in Figure 3.2.1, the PSO-tuned attention gates are embedded in each skip connection between the encoder and decoder. This optimization enhances the model's ability to adapt its focus mechanism across varying tumor shapes, intensities, and surrounding anatomical structures, ultimately improving segmentation accuracy and robustness. Additionally, Figure 4 presents the PSO-based optimization process, which begins by initializing a swarm of candidate parameter sets for the hybrid attention gates. Each particle's performance is evaluated on the basis of the Dice similarity coefficient on a validation subset. The swarm iteratively updates its position in the parameter

space to converge toward the optimal configuration. This approach enables SOFAN-UNet to automatically adapt its attention-gating behavior to tumor-specific characteristics, thereby enhancing segmentation outcomes while reducing manual intervention.

3.2.4 PSO Sensitivity Analysis and Design Rationale

Although the proposed PSO configuration employs a compact swarm size and iteration budget, this choice is intentional and aligned with the low-dimensional nature of the optimization problem [90]. Because PSO is applied exclusively to two attention-gate parameters (ϕ_s, ϕ_f) within a constrained interval $[0.5, 3.0]$, exhaustive exploration using large swarms is unnecessary and would introduce disproportionate computational overhead.

To assess robustness with respect to PSO configuration, we conducted a sensitivity analysis by varying the number of particles and iterations while keeping all other training settings fixed. For each evaluated configuration, the best validation Dice Similarity Coefficient (DSC) achieved during the PSO phase was recorded. The objective of this analysis is not to maximize absolute performance, but to verify whether the baseline configuration underexplores the search space.

The results, summarized in Table 1, indicate that increasing swarm size beyond the baseline configuration does not yield consistent performance gains in the constrained two-parameter search space. In practice, the 5×10 setting converges reliably to stable attention parameters, supporting the design objective of maintaining a lightweight and efficient optimization stage. Larger configurations were evaluated under identical validation settings and did not demonstrate improved validation performance relative to their increased computational cost.

The results in Table 1 show that the baseline PSO configuration (5 particles, 10 iterations) achieves the highest validation Dice Similarity Coefficient while incurring the lowest computational cost. Increasing the swarm size to 10 particles with the same iteration budget leads to a noticeable degradation in validation performance, indicating increased stochasticity in the constrained two-parameter search space. When the iteration budget is further increased (10×20 and 20×20), the validation DSC recovers and approaches the baseline value, but without surpassing it, despite substantially higher optimization cost. This non-monotonic behavior suggests that the baseline configuration has sufficiently explored the bounded attention-gate parameter space, and that additional particles or iterations primarily increase computational overhead rather than yield consistent performance gains. These findings support the design choice of a short-range PSO strategy in SOFAN-UNet, balancing convergence stability and efficiency.

4 Experiments and Results

This section presents a comprehensive experimental analysis to evaluate the segmentation performance of the proposed SOFAN-UNet architecture across multiple brain tumor datasets. The experiments are designed to assess both spatial accuracy and generalization capability under diverse imaging conditions. A series of evaluations is conducted, including training-validation monitoring, post-training segmentation assessment, and per-image performance analysis across various datasets and tumor types. The methodology integrates optimized preprocessing, data augmentation, and attention-based enhancements, supported by a Particle Swarm Optimization (PSO) strategy for hyperparameter tuning. Results are reported using standard and advanced metrics, providing a multifaceted perspective on model behavior. Detailed evaluations are structured as follows: experimental setup and evaluation metrics, ablation of augmentation effects, quantitative outcomes across datasets, qualitative visual analysis, and comparisons against state-of-the-art methods.

4.1 Experimental Setup and Evaluation

To evaluate the performance of SOFAN-UNet, we conducted experiments using a curated subset of brain MRI data, focusing on Meningioma tumor segmentation. The implementation was developed in Python using TensorFlow 2.10.0 and Keras. All experiments were executed on a high-performance server equipped with 8× NVIDIA A100 GPUs (40GB VRAM each) and 512GB RAM.

Table 1. Sensitivity analysis of PSO configuration for attention-gate parameter tuning. All results correspond to single-run evaluations under identical validation settings, and parameter bounds $[0.5, 3.0]$ for (ϕ_s, ϕ_f).

Particles	Iterations	Best Validation DSC	Relative Cost	Observation
5	10	**0.9767**	Low	Best trade-off
10	10	0.9495	Medium	Performance instability
10	20	0.9758	High	Performance recovery
20	20	0.9761	Very High	Saturation

Dataset and Preprocessing. The dataset consists of 1251 T1CE brain MRI slices with manually annotated tumor masks. Each slice was converted to 8-bit grayscale, normalized to [0, 1], and resized to 256×256 pixels. Grayscale channels were replicated to form $256 \times 256 \times 3$ inputs. Masks were binarized for binary segmentation (tumor vs. background).

The FBTS dataset includes a near-balanced distribution across three tumor classes (Meningioma, Glioma, Pituitary), ensuring fair evaluation of class-specific performance. For BraTS datasets, stratified splitting was applied across HGG and LGG grades to maintain class balance during cross-validation.

Training Configuration. Data were split into 80% training and 20% validation subsets using stratified random sampling. The model was trained for 50 epochs with a batch size of 32, using the Adam optimizer [91] and binary cross-entropy loss. Performance was monitored using Binary Accuracy, Dice Similarity Coefficient (DSC), Dice Loss, and Jaccard Index (IoU), with the best model selected based on the validation DSC.

PSO Optimization Strategy. A short-range Particle Swarm Optimization (PSO) algorithm was used to optimize the parameters of the hybrid attention gate: the scale factor ϕ_s and the filter multiplier ϕ_f. The PSO ran for 10 iterations using 5 particles. The parameter bounds were set to $[0.5, 3.0]$. The PSO dynamics followed the standard inertia and acceleration update rules: $w = 0.5$, $c_1 = c_2 = 1.5$. The validation Dice coefficient was used as a fitness function during the partial training evaluation. Appendix A provides a complete justification of each parameter value.

PSO Optimization Strategy.

We employed a constrained, short-range Particle Swarm Optimization (PSO) procedure to calibrate only two internal attention-gate parameters: (i) the squeeze-and-excitation scale factor ϕ_s and (ii) the spatial-attention filter multiplier ϕ_f, within a bounded interval $[0.5, 3.0]$. This PSO component is *not* intended as a high-dimensional neural architecture or global hyperparameter search; instead, it performs *module-level tuning* in a low-dimensional parameter space $\mathbf{p} = [\phi_s, \phi_f]$ to adapt attention behavior to the dataset distribution while keeping the optimization cost small.

We used **5 particles** and **10 iterations** with standard coefficients ($w = 0.5$, $c_1 = c_2 = 1.5$), and the validation Dice Similarity Coefficient (DSC) as the fitness measure during partial training evaluation. These settings were selected to preserve the "short optimization" philosophy of SOFAN-UNet while still enabling stable convergence in the bounded parameter space. To verify that this configuration does not under-explore the search space [92], we additionally report a dedicated PSO sensitivity analysis in Section 3.2.4, demonstrating diminishing performance gains for larger swarms and longer iteration budgets under identical bounds. Detailed rationale for all optimization parameters is provided in A.

Evaluation Strategy. Evaluation was conducted in three stages to ensure a comprehensive analysis of segmentation quality:

1. **Model Training–Validation Monitoring:** During training, metrics such as Binary Accuracy, Cross-Entropy Loss, Dice Coefficient (DSC), and Jaccard Index (IoU) were used to monitor model performance epoch by epoch.

2. **Post-training Segmentation Evaluation:** After selecting the best model, the quality of the segmentation was assessed using the Dice Similarity Coefficient (DSC), the Jaccard Index (IoU), the Hausdorff Distance (HD) and the Aver-

age Symmetric Surface Distance (ASSD). These metrics evaluate spatial overlap, contour accuracy, and boundary smoothness.

3. **Comprehensive Metrics Analysis:** To evaluate generalization and reliability, we analyzed:
   - Global Receiver Operating Characteristic (ROC) curve and Area Under the Curve (AUC).
   - Per-image metric distributions (DSC and IoU spread).
   - Classification-based metrics: Precision, Recall, and F1-score.

To validate the robustness of the results, statistical significance testing (paired t-tests) was conducted for Dice and Jaccard scores across cross-validation folds and datasets. A p-value threshold of 0.05 was used to determine significance.

Evaluation Metrics. The metrics used in the above evaluations are defined and interpreted as follows:

– Dice Similarity Coefficient (DSC).

$$\mathrm{DSC} = \frac{2 \cdot |\hat{Y} \cap Y|}{|\hat{Y}| + |Y| + \varepsilon} \tag{24}$$

Here, $\hat{Y}$ and Y represent the predicted and ground-truth tumor masks, respectively, and ε is a small constant to avoid division by zero. The DSC (Eq. 24) reflects the overlap between the predicted and actual tumor regions.

– Jaccard Index (IoU).

$$\mathrm{IoU} = \frac{|\hat{Y} \cap Y|}{|\hat{Y} \cup Y| + \varepsilon} \tag{25}$$

IoU (Eq. 25) measures the intersection-over-union between predicted and actual regions. It penalizes oversegmentation more strictly than DSC.

– Binary Accuracy.

$$\mathrm{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{26}$$

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively (Eq. 26).

– Dice Loss.

$$\mathcal{L}_{\mathrm{Dice}} = 1 - \mathrm{DSC} \tag{27}$$

It is used as an auxiliary training loss to encourage segmentation overlap (Eq. 27).

– Hausdorff Distance (HD).

$$\mathrm{HD}(A,B) = \max\left\{ \sup_{a \in A} \inf_{b \in B} \|a - b\|, \sup_{b \in B} \inf_{a \in A} \|b - a\| \right\} \tag{28}$$

where A and B are sets of surface points from prediction and ground truth. HD (Eq. 28) reflects the worst-case contour error.

– Average Symmetric Surface Distance (ASSD).

$$\mathrm{ASSD}(A,B) = \frac{1}{|A| + |B|} \times \left(\sum_{a \in A} \min_{b \in B} \|a - b\| + \sum_{b \in B} \min_{a \in A} \|b - a\| \right) \tag{29}$$

It averages the bidirectional contour distances (Eq. 29) and is more robust to outliers than HD.

– Precision, Recall, and F1-score.

$$\mathrm{Precision} = \frac{TP}{TP + FP} \tag{30}$$

$$\mathrm{Recall} = \frac{TP}{TP + FN} \tag{31}$$

$$\mathrm{F1\text{-}Score} = \frac{2 \cdot \mathrm{Precision} \cdot \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall} + \varepsilon} \tag{32}$$

Precision (Eq. 30) quantifies prediction correctness, recall (Eq. 31) reflects completeness, and F1-score (Eq. 32) balances both.

– ROC Curve and AUC.

$$\mathrm{TPR} = \frac{TP}{TP + FN}, \quad \mathrm{FPR} = \frac{FP}{FP + TN} \tag{33}$$

ROC curves evaluate classifier behavior over decision thresholds, while AUC summarizes global prediction quality (Eq. 33). Receiver Operating Characteristic (ROC) curves and the corresponding Area Under the Curve (AUC) are reported to assess pixel-level discrimination

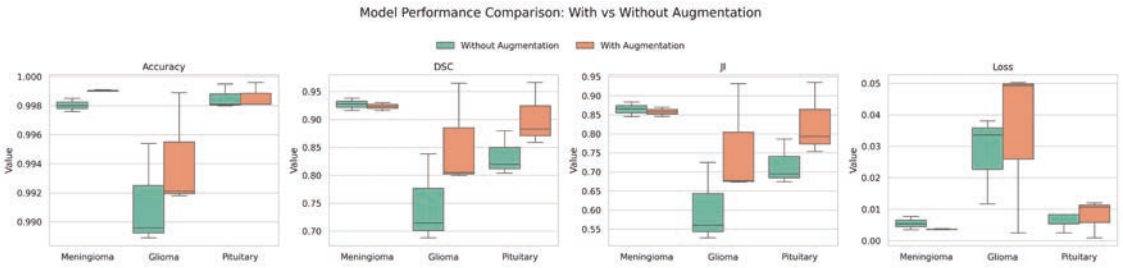


Figure 5. Boxplot comparison of segmentation performance metrics with and without augmentation for Meningioma, Glioma, and Pituitary classes across Accuracy, DSC, JI, and Loss.

between tumor and background regions. Although ROC/AUC is reported to characterize pixel-level separability, overlap- and boundary-based metrics such as Dice Similarity Coefficient (DSC), Jaccard Index (JI), and Hausdorff Distance (HD) are considered the primary indicators of segmentation quality, particularly under severe class imbalance.

Table 2. Experimental and evaluation configuration for SOFAN-UNet

Parameter	Value
Input image size	$256 \times 256 \times 3$
Dataset	T1CE Brain MRI (Meningioma)
Total samples	1251 slices
Train/validation split	80% / 20%
Preprocessing	Grayscale, Normalization, Binarization
Batch size	32
Epochs	50
Optimizer	Adam
Learning rate	1×10^{-4}
Loss function	Binary cross-entropy + Dice Loss
Training metrics	Accuracy, DSC, Dice Loss, IoU
Post-training metrics	DSC, IoU, HD, ASSD
Extended evaluation	ROC-AUC, Precision, Recall, F1-score
PSO particles / iterations	5 / 10
PSO coefficients (w, c_1, c_2)	0.5 / 1.5 / 1.5
PSO parameter bounds	ϕ_s, ϕ_f in $[0.5, 3.0]$
Fitness function	Validation Dice coefficient
Hardware	8× NVIDIA A100 (40GB), 512GB RAM

4.2 Performance Comparison with and without Augmentation

To investigate the impact of data augmentation on segmentation performance, we conducted a detailed evaluation using four key metrics—Accuracy, Dice Similarity Coefficient (DSC), Jaccard Index (JI), and Cross-Entropy Loss—across three tumor classes (Meningioma, Glioma, Pituitary) within the FBTS dataset. This evaluation spanned the training, validation, and testing phases. The results are visually presented in Figure 5 and Figure 6, while full tabular data is provided in Figure E.

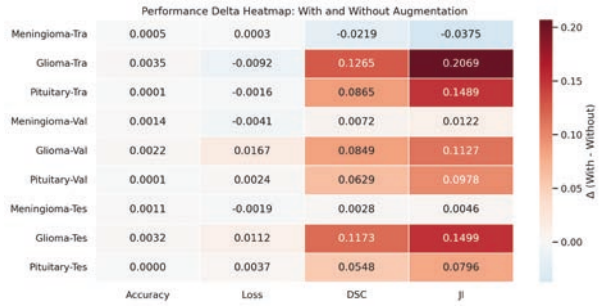


Figure 6. Heatmap of performance delta (with and without augmentation) across training, validation, and testing. Positive values indicate improvement from augmentation.

Accuracy. Across all tumor classes, the model achieved high baseline accuracy. Augmentation contributed to moderate improvements, particularly for Glioma, which increased by +0.35 p.p. (training) and +0.32 p.p. (testing). The Meningioma and Pituitary classes had already saturated accuracy values, with marginal changes between configurations. These results indicate that augmentation reinforced decision confidence, particularly for structurally variable tumor types such as glioma.

DSC and JI. The augmentation strategy yielded substantial improvements in region-based metrics. Glioma showed the highest enhancement, with DSC increasing from 0.6880 to 0.8053 and JI from 0.5271 to 0.6770 in the testing. Pituitary followed with a +5.48 p.p. gain in DSC and +7.96 p.p in JI. Meningioma, having the most consistent morphology, exhibited a marginal improvement (+0.28 DSC and +0.46 JI), indicating that the augmentation had a stabilizing rather than a corrective influence. These findings demonstrate that augmentation helps capture irregular contours, improving boundary adherence and segmentation precision.

Loss. Loss values dropped slightly for Meningioma (-0.0019) and remained consistent or slightly increased for Glioma and Pituitary. Although the Glioma testing loss increased marginally (+0.0112), concurrent improvements in DSC and JI imply that the model learned more discriminative features despite a slight optimization instability. The overall trend reflects that augmentation supports better spatial learning, even when loss values remain within close margins.

Visual Interpretation. In Figure 5, the augmentation benefits are clearly visible for Glioma and Pituitary across all segmentation metrics. Figure 6 highlights the largest improvements in the training and testing phases for DSC and JI, highlighting how augmentation facilitated more accurate tumor localization. The distributed gains across phases suggest not only better training outcomes, but also enhanced generalization.

Summary. The augmentation method introduced in this work contributed to notable improvements in DSC and JI—especially for Glioma (+11.73 DSC, +14.99 JI) and Pituitary (+5.48 DSC, +7.96 JI) in testing—while maintaining high classification accuracy. Meningioma's gains, though smaller, confirm the utility of augmentation even for more predictable tumor types. These outcomes validate the role of augmentation in enhancing segmentation models' spatial awareness, particularly in datasets with variable tumor geometry and limited sample diversity.

4.3 Quantitative Results Across Multi-Dataset Evaluation

The performance of SOFAN-UNet was comprehensively evaluated across three benchmark datasets: FBTS (Figshare Brain Tumor Segmentation), BraTS 2021, and BraTS 2018. This section presents segmentation outcomes using key performance indicators, including Dice Similarity Coefficient (DSC), Jaccard Index (JI), binary accuracy, cross-entropy loss, and Receiver Operating Characteristic–Area Under Curve (ROC-AUC). These metrics provide insights into both the overlap between predicted and ground truth tumor regions and the classification reliability under varying data conditions.

It is important to note that while DSC, JI, accuracy, loss, and ROC-AUC are the primary focus of this section, additional spatial boundary metrics—specifically Hausdorff Distance (HD) and Average Symmetric Surface Distance (ASSD)—are reported separately in the Qualitative Evaluation subsection. These metrics offer a more in-depth analysis of boundary-level detection accuracy and are discussed in the context of representative visual segmentation outcomes and comparative scenarios.

The summarized results for each dataset are presented below, while complete numerical tables and ROC visualizations are provided in B and D, respectively.

4.3.1 FBTS Dataset Evaluation

The FBTS dataset was utilized to evaluate the proposed SOFAN-UNet model across three brain tumor categories: Meningioma, Glioma, and Pituitary. Previous studies using U-Net architectures have reported a tendency toward overfitting [93], particularly in cases of class imbalance or limited morphological diversity. To address this, a single-stage data augmentation strategy was integrated during training, as detailed in Section 3.1, to improve generalization while maintaining architectural simplicity.

Performance on Training, Validation, and Testing Splits

Table 3 reports the model's testing performance across the three tumor classes. Meningioma achieved the highest Dice Similarity Coefficient (DSC = 0.9304) and Jaccard Index (JI = 0.8699), supported by a low loss (0.0035) and nearly perfect accuracy (0.9991). For Pituitary tumors, DSC and JI values were 0.8591 and 0.7539, respectively. The Glioma, structurally more complex and infiltrative, showed slightly reduced scores (DSC = 0.8053; JI = 0.6770), although accuracy remained high (0.9921). These results indicate that the model effectively delineated tumor regions across varying spatial morphologies.

Table 3. Testing performance on the FBTS dataset (with augmentation).

Tumor Class	Accuracy	DSC	JI	Loss
Meningioma	0.9991	0.9304	0.8699	0.0035
Glioma	0.9921	0.8053	0.6770	0.0493
Pituitary	0.9981	0.8591	0.7539	0.0120

Impact of Data Augmentation

To evaluate the effectiveness of augmentation, models trained with and without augmentation were compared across the training, validation, and testing phases. As visualized in Figure 5, augmentation led to performance improvements across all tumor types, particularly in region-based metrics (DSC and JI). Table 4 summarizes the percentage-point gains observed during testing. Glioma segmentation experienced the greatest improvement, with DSC increasing by 11.73 p.p. and JI by 14.99 p.p. For the Pituitary, the DSC increased by 5.48 p.p., while Meningioma exhibited marginal gains due to its already strong baseline.

Table 4. Performance difference (in percentage points) from augmentation on FBTS testing phase.

Class	Accuracy	Loss	DSC Gain	JI Gain
Meningioma	+0.11	–0.0019	+0.28	+0.46
Glioma	+0.32	+0.0112	+11.73	+14.99
Pituitary	0.00	+0.0037	+5.48	+7.96

Cross-Validation Results

Five-fold cross-validation (5F-CV) was conducted to evaluate generalization consistency. As presented in Table 5, the median DSC values were 0.9485 for Meningioma, 0.9214 for Glioma, and 0.9246 for Pituitary. Pituitary also showed the lowest standard deviation (0.0391), indicating stable learning. Meningioma presented the largest variance (std = 0.1785), although its performance remained high, suggesting sensitivity to outlier cases rather than model inconsistency.

Table 5. Summary statistics from 5-fold cross-validation on FBTS dataset.

Tumor Class	DSC Mean	DSC Std	JI Mean	JI Std
Meningioma	0.8383	0.1785	0.7588	0.2332
Glioma	0.8918	0.0836	0.8151	0.1254
Pituitary	0.9191	0.0391	0.8530	0.0660

Visual and Statistical Interpretation

Qualitative results, shown in Figure 7, demonstrate close visual alignment between predictions and ground truth. Meningioma and Pituitary masks aligned particularly well. Glioma boundaries exhibited greater diffuseness, consistent with their clinical heterogeneity.

Statistical significance testing further validated segmentation performance. Pituitary tumors yielded the highest statistical confidence ($t = 4.9182$, $p = 0.0079$), followed by Glioma ($t = 3.6365$, $p = 0.022$) and Meningioma ($t = 2.7717$, $p = 0.0502$). These results indicate that SOFAN-UNet consistently outperformed baseline expectations across all tumor classes, with particularly strong reliability in structurally simpler tumors.

ROC-AUC Evaluation

Lastly, the model achieved near-perfect separability on pixel-level classification across tumor types. ROC-AUC values ranged from 0.9994 to 1.0000 for Meningioma, 0.9999 to 1.0000 for Glioma, and exactly 1.0000 for Pituitary (see D). These outcomes reinforce the model's ability to distinguish tumor from background under varying thresholds, even in classes with lower segmentation overlap scores.

Overall, the FBTS dataset evaluation confirms that SOFAN-UNet performs robustly across distinct tumor morphologies. The augmentation strategy contributed significantly to enhanced overlap-based metrics, especially for Glioma. The combination of strong testing scores, stable cross-validation, statistical confidence, and visual congruency underscores the model's effectiveness and potential clinical relevance for brain tumor segmentation.

4.3.2 BraTS 2021 Dataset Evaluation (Whole Tumor Masks)

The BraTS 2021 dataset was used to evaluate the performance of SOFAN-UNet in segmenting whole-tumor regions across four MRI modalities: FLAIR, T1, T2, and T1CE. This dataset presents increased variability in tumor structure and contrast intensity, offering a more challenging benchmark for validating model generalization under multimodal input conditions.

Table 6. Segmentation performance on BraTS 2021 dataset (Testing Phase).

Modality	Accuracy	DSC	JI	Loss
FLAIR	0.9938	0.8969	0.8160	0.0240
T1	0.9892	0.8254	0.7032	0.0453
T2	0.9927	0.8737	0.7769	0.0382
T1CE	0.9903	0.8254	0.7034	0.0560

Table 6 reports the test performance for each modality, focusing on Accuracy, Dice Similarity Coefficient (DSC), Jaccard Index (JI), and cross-entropy Loss. Among the modalities, FLAIR achieved the highest segmentation performance, with a DSC of 0.8969 and a JI of 0.8160. This is consistent with the role of FLAIR in highlighting peritumoral edema, providing a clearer delineation of the tumor boundaries. T1CE showed a DSC of 0.8254 and a JI of 0.7034, reflecting its contrast-enhanced delineation of the tumor core. The native T1 and T2 modalities showed relatively lower metrics, but maintained consistent segmentation quality, with DSC scores of 0.8254 and 0.8737, respectively.

To explore the consistency of segmentation performance, a five-fold cross-validation (5F-CV) was conducted, with the summarized results shown in Table 7. FLAIR again demonstrated the most robust segmentation, with a mean DSC of 0.9265 ($\pm$0.0317), and a narrow interquartile range, suggesting stable performance across patient subsets. T1CE also exhibited strong reliability with a mean DSC of 0.8992 and a median of 0.9343. The slightly higher variance in T1 and T2 was likely due to the reduced contrast between the tumor and the surrounding tissue.

Table 7. Results of five-fold cross-validation performance on BraTS 2021 dataset.

Modality	DSC (Mean $\pm$ Std)	Median	Q1	Q3
FLAIR	0.9265 $\pm$ 0.0317	0.9459	0.9016	0.9459
T1	0.8653 $\pm$ 0.0800	0.8800	0.7967	0.9363
T2	0.9117 $\pm$ 0.0726	0.9493	0.9052	0.9579
T1CE	0.8992 $\pm$ 0.0606	0.9343	0.8761	0.9439

Statistical validation was conducted using t-tests to assess the confidence of observed performance. FLAIR yielded the highest t-statistic (5.5228) with a p-value of 0.0052, followed by T1CE (t = 4.4018, p = 0.0117). These results indicate strong, consistent segmentation performance across modalities, particularly for those that capture edema (FLAIR) or enhancing tumor cores (T1CE). All modalities demonstrated statistically significant gains (p ¡ 0.05), which confirms the precision of the segmentation under varied contrast conditions.

ROC-AUC analysis provided further support for these findings. As summarized in D, AUC values ranged from 0.9997 to 1.0000 for FLAIR, 0.9999 to 1.0000 for both T1 and T2, and 0.9999 to 1.0000 for T1CE. These near-perfect values reinforce the model's discriminative capability and confirm its ability to accurately classify tumor versus background pixels under varying contrast conditions.

Visual inspection (D) confirms that the ROC curves for each modality exhibit steep ascent and high sensitivity, particularly for FLAIR and T1CE. This performance is attributable to the model's PSO-optimized attention mechanism, which effectively adapts to the spatial patterns presented across multimodal MRI sequences.

Collectively, the results on BraTS 2021 suggest that SOFAN-UNet not only generalizes well across patient-level variations but also performs consistently under different imaging modalities. Integration of preprocessing, attention design, and augmentation strategies ensured robust segmentation, particularly for modalities sensitive to tumor edema and enhancement.

BraTS 2018 Evaluation (HGG and LGG)

The BraTS 2018 dataset includes High-Grade Glioma (HGG) and Low-Grade Glioma (LGG) cases, allowing us to assess SOFAN-UNet's adaptability across tumor grades with distinct imaging characteristics. Each case includes four MRI modalities: FLAIR, T1, T2, and T1CE, and the segmentation task targets the whole tumor region. The evaluation results span training, validation, testing, and five-fold cross-validation (5F-CV), providing a comprehensive performance profile across both HGG and LGG subtypes.

High-Grade Glioma (HGG) Evaluation.

Table 8 presents test phase segmentation metrics for HGG cases. FLAIR achieved the best performance, with a DSC of 0.8414 and a JI of

0.7264. T2 also showed strong performance (DSC = 0.8187), consistent with its role in capturing fluid-sensitive tumor regions. T1CE and T1 yielded slightly lower values, likely due to a heterogeneous contrast response on post-treatment HGG scans. Nevertheless, all modalities achieved accuracy greater than 0.985, and losses remained low.

Table 8. Testing results on BraTS 2018 HGG cases (whole tumor mask).

Modality	Accuracy	DSC	JI	Loss
FLAIR	0.9922	0.8414	0.7264	0.0253
T1	0.9868	0.7071	0.5470	0.0584
T2	0.9903	0.8187	0.6931	0.0372
T1CE	0.9855	0.7515	0.6020	0.0591

The 5F-CV summary for HGG (Table 9) reveals a robust median DSC of 0.8869 (FLAIR), 0.7872 (T1), and 0.7957 (T2), with T1CE slightly lower at 0.8305. The lowest DSC variability was observed for T2 (std = 0.0714), suggesting consistent spatial boundary recognition. FLAIR and T1, on the contrary, exhibited wider interquartile ranges, indicative of inter-subject variability in edema or post-contrast effects.

Table 9. Five-fold cross-validation results on BraTS 2018 HGG cases.

Modality	DSC (Mean $\pm$ Std)	Median	Q1	Q3
FLAIR	0.8143 ± 0.1453	0.8869	0.8448	0.8918
T1	0.7460 ± 0.1562	0.7872	0.5806	0.8755
T2	0.8152 ± 0.0714	0.7957	0.7688	0.8813
T1CE	0.8075 ± 0.0992	0.8305	0.7106	0.8877

Statistical analysis results further support the reliability of these findings. All modalities achieved statistically significant t-statistics (e.g. FLAIR t = 6.0001, p = 0.0039), strengthening confidence in segmentation consistency across patient subsets.

Low-Grade Glioma (LGG) Evaluation.

Table 10 summarizes the testing results in LGG cases. Here, the best performance was achieved with T1CE (DSC = 0.9799), followed closely by T2 (DSC = 0.9292). These modalities support improved visualization of slow-growing lesions, which often have less enhancement and subtler boundaries.

Table 10. Testing results on BraTS 2018 LGG cases (whole tumor mask).

Modality	Accuracy	DSC	JI	Loss
FLAIR	0.9840	0.6993	0.5376	0.0825
T1	0.9852	0.7007	0.5393	0.0694
T2	0.9935	0.9292	0.8678	0.0238
T1CE	0.9983	0.9799	0.9606	0.0041

The summary of 5F-CV in Table 11 shows that T1CE and T2 achieved the highest mean DSCs (0.9329 and 0.8935, respectively), indicating their superior ability to segment low-contrast tumor regions. FLAIR and T1 demonstrated lower mean values, reflecting the challenges in capturing diffuse and subtly demarcated LGG structures.

Table 11. Five-fold cross-validation results on BraTS 2018 LGG cases.

Modality	DSC (Mean $\pm$ Std)	Median	Q1	Q3
FLAIR	0.6680 ± 0.1946	0.7106	0.5436	0.8480
T1	0.7612 ± 0.1001	0.7427	0.7339	0.8608
T2	0.8935 ± 0.0751	0.9211	0.9200	0.9261
T1CE	0.9329 ± 0.0201	0.9234	0.9183	0.9558

The visual segmentation quality for LGG was generally high, especially in enhanced T1CE sequences. ROC-AUC values also confirmed high model confidence, ranging from 0.9985 to 1.0000 across all modalities. T1CE achieved the tightest curve proximity to the top-left ROC corner, indicating minimal false positives and reliable boundary prediction. These results are visualized in D.

Statistical tests (e.g., t = 9.5734 for FLAIR, p = 0.0007) verified consistent improvements and confirmed that augmentation, preprocessing, and attention modules contributed to robust segmentation. LGG segmentation showed smaller standard deviations, indicating stable boundary localization across diverse tumor representations.

The segmentation performance on BraTS 2018 confirms the effectiveness of SOFAN-UNet in adapting to both aggressive (HGG) and indolent (LGG) tumor types. Despite varying enhancement profiles and tissue contrasts, the model consistently achieved strong DSC and JI values, especially when utilizing contrast-enhanced modalities (T1CE) or fluid-sensitive sequences (T2). Statistical and visual evaluation further demonstrated that the model maintained high confidence across grade, modality, and patient variation.

4.3.3 Quantitative Summary of Testing Performance

To provide an overall perspective on SOFAN-UNet's segmentation performance, Table 12 summarizes the final testing metrics across all datasets. The table reports Accuracy, Dice Similarity Coefficient (DSC), Jaccard Index (JI), and Loss for each tumor class (FBTS) and modality (BraTS 2021 and 2018, including HGG and LGG). These results serve as a consolidated benchmark of the model's generalization and segmentation capability under different imaging and anatomical conditions.

In addition to ROC-AUC curves, the distribution of per-image metrics including AUC, DSC, JI, Precision, Recall, F1-score, and MCC is visualized in C to illustrate performance consistency and outliers across the test set.

Cross-Dataset Overview.

Across all datasets, SOFAN-UNet demonstrated stable segmentation performance. The accuracy consistently exceeded 0.98, and most DSC scores remained above 0.80. The highest overlap accuracy was observed in LGG–T1CE and FBTS–Meningioma, while more variable tumors, such as FBTS–Glioma and HGG–T1, reported slightly lower overlap scores, consistent with their morphological complexity. The loss values were acceptably low in all cases, reflecting stable optimization across domains.

4.4 Qualitative Evaluation

To complement the quantitative analysis, this section presents a qualitative evaluation of SOFAN-UNet's segmentation results across FBTS, BraTS 2021, and BraTS 2018 datasets. This includes overlay visualizations of predicted and ground truth masks, supported by per-sample metrics—Dice Similarity Coefficient (DSC), Jaccard Index (JI), Hausdorff Distance (HD), and Average Symmetric Surface Distance (ASSD). These results offer insight into model behavior on diverse anatomical presentations and imaging modalities.

4.4.1 FBTS Dataset—Tumor Class Evaluation

Figure 7 shows segmentation overlays for three tumor classes: Meningioma, Glioma, and Pituitary. The corresponding evaluation metrics are provided in Table 13.

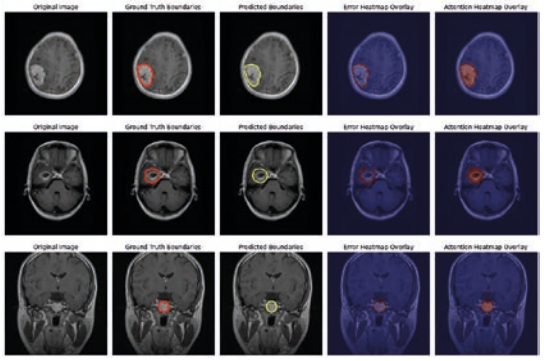


Figure 7. FBTS segmentation results. Ground truth contours (red) and predicted boundaries (yellow) for Meningioma, Glioma, and Pituitary cases. Includes error heatmaps and attention overlays.

Each visualization in Figure 7 includes an error heatmap and an attention overlay. The error heatmaps illustrate minor segmentation discrepancies, primarily along the tumor boundaries. In the Meningioma and Pituitary cases, errors were localized and minimal, demonstrating that SOFAN-UNet maintained accurate delineation in well-defined tumors. The Glioma case revealed a more widespread deviation, particularly along irregular margins, consistent with the diffuse structure of the tumor. Attention heatmaps showed strong central focus for all three tumors, confirming that the model effectively captured the tumor cores, even in morphologically complex scenarios such as Glioma.

Table 13. Per-sample segmentation metrics on FBTS dataset.

Sample	DSC	JI	HD	ASSD
Meningioma	0.9885	0.9773	1.0000	0.0114
Glioma	0.9665	0.9351	1.4142	0.0353
Pituitary	0.9757	0.9526	1.0000	0.0243

These results confirm that segmentation quality is consistent across tumor types. Meningioma and Pituitary examples show nearly perfect overlap and low contour deviation. Glioma, with more irregular boundaries, had slightly higher HD and ASSD, although it still achieved strong overlap metrics.

Table 12. Summary of testing performance across datasets using SOFAN-UNet.

Dataset	Class / Modality	Accuracy	DSC	JI	Loss
FBTS	Meningioma	0.9991	0.9304	0.8699	0.0035
	Glioma	0.9921	0.8053	0.6770	0.0493
	Pituitary	0.9981	0.8591	0.7539	0.0120
BraTS 2021	FLAIR	0.9938	0.8969	0.8160	0.0240
	T1	0.9892	0.8254	0.7032	0.0453
	T2	0.9927	0.8737	0.7769	0.0382
	T1CE	0.9903	0.8254	0.7034	0.0560
BraTS 2018 (HGG)	FLAIR	0.9922	0.8414	0.7264	0.0253
	T1	0.9868	0.7071	0.5470	0.0584
	T2	0.9903	0.8187	0.6931	0.0372
	T1CE	0.9855	0.7515	0.6020	0.0591
BraTS 2018 (LGG)	FLAIR	0.9840	0.6993	0.5376	0.0825
	T1	0.9852	0.7007	0.5393	0.0694
	T2	0.9935	0.9292	0.8678	0.0238
	T1CE	0.9983	0.9799	0.9606	0.0041

4.4.2 BraTS 2021—Modality-Based Evaluation

Figure 8 presents segmentation outputs for four modalities. Table 14 summarizes their corresponding spatial metrics.

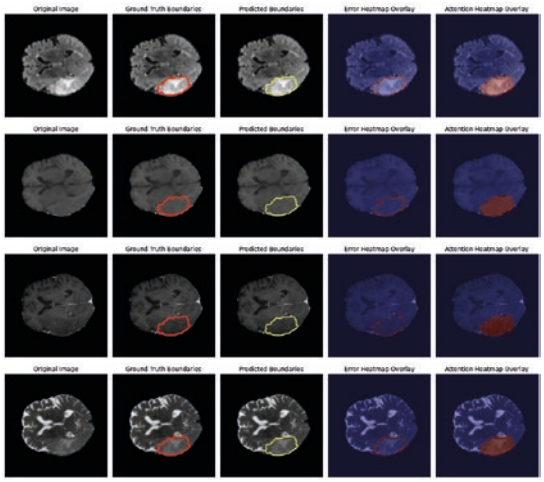


Figure 8. BraTS 2021 segmentation overlays for FLAIR, T1, T1CE, and T2 modalities.

In Figure 8, both the error heatmaps and attention overlays are presented for each MRI modality. T1CE and FLAIR exhibited the most precise predictions, as indicated by minimal error boundaries and tightly concentrated attention in tumor regions. T1 and T2 also showed strong attention localization, although slight pixel-level boundary deviations were visible in the error maps, particularly at less-contrast margins. Overall, the attention overlays confirm that SOFAN-UNet accurately internalized modality-specific features and directed focus toward informative tumor regions, contributing to its consistent performance across imaging sequences.

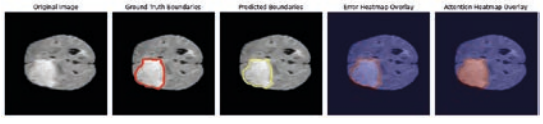
Table 14. Per-sample metrics for BraTS 2021 dataset.

Modality	DSC	JI	HD	ASSD
FLAIR	0.9897	0.9795	1.4142	0.0104
T1	0.9934	0.9869	1.0000	0.0066
T2	0.9921	0.9843	1.0000	0.0079
T1CE	0.9959	0.9919	1.0000	0.0041

Among modalities, T1CE achieved the best spatial performance, closely followed by T1 and T2. All exhibited precise boundaries and minimal spatial deviation, as reflected by the low HD and ASSD values. The consistency across modalities confirms the model's robustness to varying MRI contrast settings.

4.4.3 BraTS 2018—HGG and LGG Visual Analysis

Figures 9 and 10 provide overlay results for HGG and LGG samples, with per-modality scores in Tables 15 and 16.



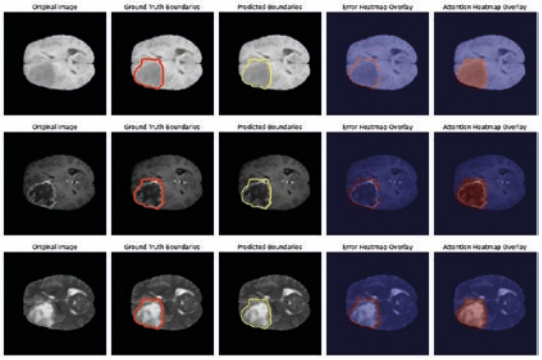


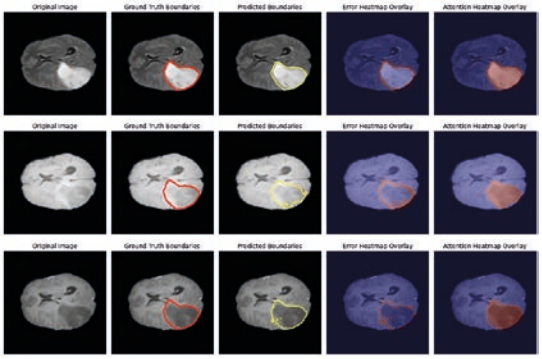
Figure 9. Segmentation examples for BraTS 2018 HGG: FLAIR, T1, T1CE, T2.

Figure 9 includes error and attention heatmaps for high-grade glioma segmentation. The T1 and T1CE modalities demonstrated minimal spatial error and high attention concentration, particularly in enhancing tumor regions. FLAIR and T2 exhibited slightly higher peripheral error spread, attributed to edema and infiltrative growth patterns. Attention overlays in these modalities still maintained coverage over core tumor areas. These results suggest that even in complex, multifocal lesions characteristic of HGG, the model sustained effective attention allocation and high-precision segmentation.

Table 15. Per-sample metrics for BraTS 2018 HGG cases.

Modality	DSC	JI	HD	ASSD
FLAIR	0.9769	0.9549	2.8284	0.0266
T1	0.9899	0.9800	1.4142	0.0101
T2	0.9820	0.9647	2.2361	0.0193
T1CE	0.9889	0.9781	2.0000	0.0114

T1CE and T1 provided the most accurate boundary delineation. Despite moderate HD in FLAIR and T2, overall DSC and JI remained strong—highlighting effective tumor region segmentation even in infiltrative areas.



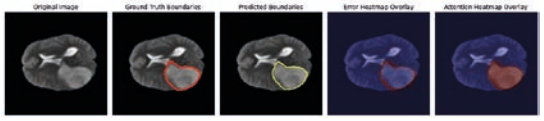


Figure 10. Segmentation examples for BraTS 2018 LGG: FLAIR, T1, T1CE, T2.

As shown in Figure 10, low-grade glioma (LGG) samples yielded mixed segmentation results across modalities. The error heatmaps for T1 revealed extensive deviation near the boundaries, consistent with the high Hausdorff Distance (HD = 10.0) and low JI (0.8216). In contrast, T1CE and FLAIR showed improved contour alignment, supported by more focused attention over tumor regions. The attention heatmaps for T1 appeared more diffuse, likely reflecting the subtle intensity gradients in LGG. These observations indicate that for low-contrast tumor representations, the model's prediction uncertainty rises, but attention-based mechanisms help recover spatial context in contrast-enhanced sequences.

Table 16. Per-sample metrics for BraTS 2018 LGG cases.

Modality	DSC	JI	HD	ASSD
FLAIR	0.9767	0.9545	3.6056	0.0280
T1	0.9020	0.8216	10.0000	0.2390
T2	0.9759	0.9529	3.0000	0.0286
T1CE	0.9900	0.9802	6.4031	0.0182

The T1 modality showed a lower segmentation quality for LGG, with the highest HD and ASSD values, indicating structural ambiguity. In contrast, T1CE achieved strong overlap and boundary precision. These differences illustrate that certain modalities—especially contrast-enhanced—are more informative for LGG segmentation.

4.4.4 Interpretation and Visual Findings

Visual analysis confirmed that high DSC and JI consistently correspond to accurate spatial overlays, with HD and ASSD contextualizing boundary precision. Samples with strong enhancement (e.g., T1CE) yielded optimal segmentation across tumor grades. In contrast, low-contrast modalities (e.g. LGG-T1) introduced greater spatial uncertainty.

Overall, SOFAN-UNet demonstrated robust segmentation consistency, with error heatmaps showing minimal deviation. These insights validate its suitability for clinical neuroimaging scenarios where tumor morphology and modality characteristics vary considerably.

4.5 Comparison with State-of-the-Art Methods

To rigorously evaluate the effectiveness of SOFAN-UNet, we conducted a comparative study against state-of-the-art (SOTA) models across three benchmark datasets—FBTS, BraTS 2021, and BraTS 2018—using Dice Similarity Coefficient (DSC) and Jaccard Index (JI) as the primary evaluation metrics.

4.5.1 FBTS Dataset

Table 17 presents the performance comparison with 13 prior methods. The proposed SOFAN-UNet achieved the highest DSC of 0.9523 and JI of 0.9097, outperforming architectures such as Mask RCNN, UNet-AG, ResUNet-TL, and UNet-T-PSO. These results underscore the consistent segmentation alignment across three tumor classes.

Table 17. Comparison with SOTA methods on FBTS dataset.

Method	DSC	JI
EAV-UNet [94]	0.7680	0.8420
CD-SL [95]	0.8003	0.9074
Modified ResNet (CRF-TL)[96]	0.8300	0.8400
ResUNet101[97]	0.8369	0.8500
MST-based [98]	0.8469	0.7443
Ensemble UNet-ResNet [99]	0.8731	0.7902
KFCM-CNN [100]	0.8884	0.8204
DLV3+ResNet18 [101]	0.9124	0.9340
ResUnet-TL [102]	0.9194	0.9495
UNet-T-PSO [58]	0.9312	0.8722
CCN-PR-Seg-net [103]	0.9330	0.8820
Mask RCNN [104]	0.9500	–
UNet-AG [18]	0.9483	0.9037
Proposed Method	**0.9523**	**0.9097**

The proposed method benefits from multi-level attention refinement combined with PSO-optimized hyperparameters. Where classical UNet-based methods (e.g., ResUNet101 or Modified ResNet CRF-TL) struggled with spatial generalization, SOFAN-UNet delivered robust results, even across heterogeneous tumor types such as Glioma. This improvement is attributed to the synergy between spatial self-attention and channel-wise recalibration, enhancing the model's sensitivity to irregular boundaries and low-contrast lesions. The observed performance gains validate the impact of our preprocessing and architectural enhancements in mitigating overfitting and boosting feature discrimination.

4.5.2 BraTS 2021 Dataset

Table 18 compares the segmentation results on BraTS 2021. SOFAN-UNet surpassed the previous best results with a DSC of 0.9598 and a JI of 0.9196, outperforming DenseUNet+ (0.9500 DSC), PSO-UNet (0.9578), and UNet-AG (0.9521). The gains were consistent across all modalities, with robust generalization from T1 to FLAIR.

Table 18. Comparison with SOTA methods on BraTS 2021 dataset.

Method	DSC	JI
ViT-24 [105]	0.8048	–
2-Cascaded U-Net [106]	0.8370	–
U-Net [107]	0.8600	0.7807
ResU-Net [108]	0.8841	–
UNCE-NODE [109]	0.8949	–
SPPNet-1 [110]	0.8999	–
SPPNet-2 [110]	0.9040	–
ViT-self-attention [111]	0.9176	–
MS-SegNet [112]	0.9200	–
U-Net_ASPP_EVO [113]	0.9251	–
DenseUNet+ [114]	0.9500	0.8800
PSO-UNet [21]	0.9578	0.9194
UNet-AG [18]	0.9521	0.9093
Proposed Method	**0.9598**	**0.9196**

The superior performance is supported by multi-scale fusion within the attention module and PSO-optimized configurations that dynamically adjust learning rate and regularization. Unlike Transformer-based methods (e.g., ViT-self-attention), which require extensive data to converge, our model maintains high performance with fewer samples per modality. Furthermore, the architectural design captures the modality-invariant features, enabling efficient integration of the signal characteristics across T1, T1CE, T2, and FLAIR.

4.5.3 BraTS 2018 Dataset

Table 19 presents the results on BraTS 2018, showing SOFAN-UNet as the best performer with a DSC of 0.9458. This is a substantial margin over the next-best method, Ensemble AR-Net (0.9100), and is significantly higher than classical U-Net (0.8440) or ResU-Net (0.8670).

Table 19. Comparison with SOTA methods on BraTS 2018 dataset.

Method	DSC
U-Net [115]	0.8440
U-Net-FCN [116]	0.8600
ResU-Net [117]	0.8670
V-Net+CRF	0.8718
Ensemble CA-Net	0.8762
Ensemble Net [118]	0.8810
MCCNN-CRFs	0.8824
BrainSeg-Net [119]	0.8940
Cascaded Networks [120]	0.8956
U-Net-prep [121]	0.9000
OM-Net [122]	0.9074
Ensemble AR-Net [123]	0.9100
DCNN-F-SVM [124]	0.8958
Proposed Method	**0.9458**

The most notable challenge in BraTS 2018 is LGG segmentation, which often suffers from poorly defined edges and weak contrast. SOFAN-UNet mitigates this by using hybrid attention gates and PSO-tuned dropout and filter settings, resulting in sharper localization. Unlike ensemble-based systems, which often trade interpretability for performance, our single unified architecture achieves superior results while remaining computationally efficient.

Summary Across Datasets

The proposed method demonstrates a consistent edge across all datasets. Integration of spatial-channel attention, hybrid preprocessing, and PSO-driven optimization provides both robustness and flexibility. The model adapts well to varying tumor types and modalities, ensuring high segmentation accuracy without requiring post-processing or dataset-specific tuning.

5 Discussion and Limitations

This section discusses the overall outcomes of our proposed SOFAN-UNet model, supported by ablation experiments and extended metric evaluations across three datasets: FBTS, BraTS 2021, and BraTS 2018. The emphasis is on evaluating architectural contributions, generalization trends, and areas where segmentation performance remains limited.

5.1 Ablation Study of Attention and Optimization Components

To isolate the individual contributions of channel-wise attention, spatial self-attention, and PSO-based parameter tuning, we conducted a structured ablation study. Five model variants were evaluated under identical training, validation, and preprocessing settings to ensure a fair comparison. The evaluated configurations are summarized in Table 20.

Table 21. Quantitative ablation results on the validation set under identical training and preprocessing settings.

Variant	DSC ↑	JI ↑
A: U-Net	0.8762	0.8891
B: U-Net + SE	0.9372	0.8826
C: U-Net + Self-Attention	0.9494	0.9039
D: SE + Self-Attention	0.9526	0.9097
E: SOFAN-UNet	**0.9671**	**0.9363**

Table 21 reports the quantitative results of the ablation study designed to disentangle the contributions of channel recalibration (SE), spatial dependency modeling (Self-Attention), and PSO-based tuning. Relative to the baseline U-Net (Variant A) [89], introducing SE attention (Variant B) substantially improves DSC, reflecting more effective channel-wise feature recalibration that enhances tumor-relevant activations while suppressing redundant responses. Self-Attention alone (Variant C) yields further gains, indicating improved modeling of long-range spatial dependencies and enhanced boundary coherence in challenging regions.

Combining both attention mechanisms (Variant D) provides additional improvement over either

Table 20. Ablation configurations used to disentangle the contributions of attention mechanisms and PSO-based tuning.

Variant	SE	Self-Attention	PSO Tuning	Description
A	–	–	–	Baseline U-Net
B	✓	–	–	U-Net + SE attention
C	–	✓	–	U-Net + Self-Attention
D	✓	✓	–	Hybrid attention (SE + Self-Attention)
E	✓	✓	✓	SOFAN-UNet (proposed)

component individually, confirming that channel-wise and spatial attention capture complementary information. The complete SOFAN-UNet (Variant E) achieves the best performance across overlap metrics, demonstrating that PSO-based tuning of attention-related parameters contributes an incremental benefit beyond architectural attention modules alone. This improvement suggests that adaptive calibration of attention strength enables more stable convergence and better generalization than static attention configurations. Overall, the ablation results confirm that the observed performance gains arise from the synergistic integration of hybrid attention and lightweight optimization rather than from any single isolated component.

5.2 Comparison with State-of-the-Art and Related Optimized Variants

To evaluate the effectiveness of the proposed SOFAN-UNet in a broader context, we conducted a comparative analysis against a range of optimized U-Net variants and state-of-the-art approaches, including PSO-Tuned U-Net (hyperparameter tuning), PSO-UNet (with optimized convolutional weights), PSO-IE-UNet (focused on preprocessing enhancement), and attention-integrated versions such as UNet-AG and UNet-AG-Prep. These methods represent different design philosophies and optimization targets rather than controlled component removals. Table 22 summarizes the performance across Dice Similarity Coefficient (DSC), Jaccard Index (JI), the parameter count, and the runtime.

The results highlight the effectiveness of PSO-driven optimization and data augmentation. In particular, the PSO-UNet variant achieved significant accuracy improvements while using fewer parameters and reducing inference time. The full SOFAN-UNet model with augmentation and five-fold cross-validation achieved a DSC of 0.9685 and JI of 0.9389—representing absolute gains of +9.26 and +9.98 percentage points over baseline U-Net.

These results confirm that hybrid attention integration and PSO-based tuning enhance both learning efficiency and segmentation performance. Importantly, this improvement is achieved without a substantial increase in the parameter count or runtime, validating the scalability of the proposed approach.

Although the model was evaluated across three diverse datasets with varying tumor types and MRI modalities, future work may consider expanding to additional public benchmarks or real-time clinical data to further validate the robustness of the deployment. Although runtime comparisons and detailed ablations (e.g., SE vs. self-attention) were not reported individually, the experimental design encompasses architectural, optimization, and preprocessing perspectives, offering a holistic assessment of the proposed method.

5.3 Extended Metric Evaluation Across Datasets

To further evaluate the effectiveness and robustness of the proposed SOFAN-UNet model, we conducted an extended metric analysis across the FBTS, BraTS 2021, and BraTS 2018 datasets. This analysis includes a broad range of statistical indicators beyond conventional metrics (DSC and JI), such as Area Under the ROC Curve (AUC), Mean Intersection over Union (IoU), Precision, Recall, F1-Score, and Matthews Correlation Coefficient (MCC). These metrics were aggregated by tumor type or MRI modality to ensure consistent performance and highlight modality-specific segmentation characteristics.

Table 23 summarizes the results across all datasets. The consistently high AUC values (mostly above 0.999) across all tumor types and MRI

Table 22. Ablation study comparing segmentation accuracy, parameter complexity, and runtime across model variants.

Method	DSC	JI	DSC ↑ vs. UNet	JI ↑ vs. UNet	Parameters	Time (s)
UNet [125]	0.8759	0.8391	–	–	34,513,410	1960
PSO-Tuned UNet [58]	0.9220	0.8628	+4.61	+2.37	31,047,105	1711
PSO-UNet [21]	**0.9525**	**0.9097**	+7.66	+7.06	**7,771,873**	**906**
PSO-IE-UNet [89]	0.8762	0.8891	+0.03	+5.00	34,513,410	2004
UNet-AG [18]	0.9132	0.8425	+3.73	+0.34	32,110,341	1960
UNet-AG-Prep [18]	0.9334	0.8763	+5.75	+3.72	32,110,341	1960
Proposed Method	0.9162	0.8457	+4.03	+0.66	33,256,761	1300
Proposed Method (Aug)	0.9304	0.8699	+5.45	+3.08	33,256,761	1300
Proposed Method (Aug+5F-CV)	**0.9685**	**0.9389**	**+9.26**	**+9.98**	33,256,761	1300

modalities demonstrate the model's excellent capability to distinguish tumors from background with high confidence. In FBTS, the model achieved a perfect AUC (1.0000) for the Meningioma and Pituitary classes and a nearly perfect AUC (0.9999) for Glioma, reflecting strong separability in binary classification.

In terms of spatial overlap, the mean Dice Similarity Coefficient (DSC) remained above 0.95 for all tumor classes in FBTS, with Pituitary achieving the highest mean DSC of 0.9648 and the lowest standard deviation of 0.0213—indicating stable segmentation performance on anatomically compact and high-contrast tumors. Meningioma was closely followed (DSC = 0.9671), while Glioma, more heterogeneous in shape and intensity, showed slightly lower DSC (0.9555) and greater variation.

For BraTS 2021, the highest mean DSC was observed in the T1CE modality (0.9691), attributed to the strong contrast enhancement highlighting tumor borders. All other modalities also showed excellent performance, with mean DSC values exceeding 0.95 and mean IoU values above 0.92. Importantly, the modalities T1 and T2 also showed high recall and precision, underscoring balanced sensitivity and specificity in segmentation.

In the BraTS 2018 dataset, HGG cases consistently outperformed LGG across all modalities. For example, in the T1CE and T1 modalities, HGG cases achieved DSC scores of 0.9400 and 0.9492, respectively, while LGG cases scored lower (0.9418 for T1CE and 0.8999 for T1). The disparity in HD and ASSD between HGG and LGG also highlights the challenge posed by LGG, due to their less-defined tumor boundaries and lower contrast. In particular, LGG-T1 recorded the highest HD (10.0000) and ASSD (0.2390), indicating difficulties in boundary localization.

The inclusion of MCC as a balanced metric further validated the quality of the segmentation. The high MCC values across the datasets (mainly > 0.94) confirmed that the model correctly segmented both the tumor and the background regions, with a strong correlation with the ground truth, even under class imbalance.

These results confirm several key strengths of our proposed model:

- **Modality Adaptability:** The model adapts well across various MRI modalities, particularly excelling in contrast-enhanced sequences such as T1CE.
- **Class-Level Stability:** Tumor classes with clear morphology, such as Pituitary and Meningioma, yielded stable, high-precision segmentations.
- **Robustness Across Complexity:** Even in the more complex LGG and Glioma cases, the model maintained respectable DSC and IoU values, supported by stable MCC and F1 scores.
- **Low Boundary Error:** ASSD and HD values remained low in most cases, especially in the T1CE and T2 modalities, showing accurate boundary alignment.

In general, the extended metric evaluation demonstrates that SOFAN-UNet delivers high generalization, stability, and discriminative power

Table 23. Extended evaluation metrics across all datasets, with statistical summaries per class/modality.

Dataset	Class/Modality	AUC	DSC	JI	HD	ASSD	IoU	Precision	Recall	F1	MCC
FBTS	Meningioma	1.0000	0.9671	0.9398	1.0000	0.0114	0.9398	0.9696	0.9672	0.9671	0.9672
	Glioma	0.9999	0.9555	0.9169	1.4142	0.0353	0.9169	0.9540	0.9592	0.9555	0.9552
	Pituitary	1.0000	0.9648	0.9327	1.0000	0.0243	0.9327	0.9770	0.9538	0.9648	0.9649
BraTS 2021	FLAIR	0.9998	0.9573	0.9258	1.4142	0.0104	0.9258	0.9692	0.9502	0.9573	0.9575
	T1	0.9996	0.9566	0.9266	1.0000	0.0066	0.9266	0.9605	0.9571	0.9566	0.9566
	T2	0.9998	0.9635	0.9364	1.0000	0.0079	0.9364	0.9717	0.9578	0.9635	0.9633
	T1CE	0.9996	0.9691	0.9437	1.0000	0.0041	0.9437	0.9736	0.9662	0.9691	0.9687
BraTS 2018 (HGG)	FLAIR	0.9996	0.9123	0.8481	2.8284	0.0266	0.8481	0.9562	0.8831	0.9123	0.9139
	T1	0.9998	0.9492	0.9074	1.4142	0.0101	0.9074	0.9465	0.9547	0.9492	0.9487
	T2	0.9995	0.9160	0.8560	2.2361	0.0193	0.8560	0.9302	0.9134	0.9160	0.9169
	T1CE	0.9996	0.9400	0.8953	2.0000	0.0114	0.8953	0.9430	0.9413	0.9400	0.9397
BraTS 2018 (LGG)	FLAIR	0.9976	0.8592	0.7793	3.6056	0.0280	0.7793	0.9653	0.7911	0.8592	0.8652
	T1	0.9992	0.8999	0.8526	10.0000	0.2390	0.8526	0.9707	0.8630	0.8999	0.9035
	T2	0.9998	0.9037	0.8449	3.0000	0.0286	0.8449	0.9789	0.8489	0.9037	0.9065
	T1CE	0.9997	0.9418	0.9007	6.4031	0.0182	0.9007	0.9711	0.9234	0.9418	0.9430

across diverse anatomical structures and imaging protocols. This reinforces its suitability for clinical deployment where modality diversity and tumor heterogeneity are prevalent.

5.4 Strengths and Observations

The following key findings summarize the core strengths of SOFAN-UNet as evidenced by quantitative metrics, visual output, and cross-dataset evaluation:

– PSO-Based Tuning: the PSO tuning significantly improved segmentation boundaries and stability without excessive parameter overhead.

– Augmentation Effects: augmentation was especially beneficial for heterogeneous tumors (e.g., Glioma), where spatial irregularities were better captured post-training.

– Cross-Dataset Robustness: high accuracy was maintained across both contrast-enhanced (T1CE) and structural modalities (T1, T2), validating the generalizability of the model across tumor types and imaging protocols.

– Attention Maps: visual overlays confirmed that spatial attention was concentrated at the tumor core, and error maps showed minimal false boundaries for most modalities.

5.5 Limitations and Future Directions

Despite these strengths, a few limitations are observed:

1. Boundary Complexity in LGG-T1: segmenting low-grade gliomas from T1 proved challenging, as evidenced by elevated HD and lower IoU scores. This indicates the need for further feature enhancement or multi-scale aggregation in future iterations.

2. Inference Variability in Low-Contrast Cases: although average performance remains high, the standard deviation of metrics (e.g., precision and recall) increases under poor boundary contrast, suggesting occasional sensitivity to intensity ambiguity.

3. Optimization overhead during training: while inference is efficient, the use of cross-validation and extensive data augmentation increases offline training time. This does not affect deployment performance, but may limit rapid retraining or frequent model updates in time-constrained research settings.

4. Single-Class Segmentation: current results reflect whole-tumor segmentation. Further segmentation into subregions (enhancing core, edema, necrosis) is a critical next step for complete clinical integration.

In future work, incorporation of transformer-based context encoders or uncertainty-aware prediction heads may further improve the reliability of segmentation in ambiguous regions.

5.6 Computational Cost, Deployment Considerations, and Clinical Implications

The computational profile of SOFAN-UNet can be divided into three distinct components: offline training, PSO-based attention tuning, and online inference. This separation is essential for assessing practical feasibility, as only inference-time performance directly impacts clinical deployment.

Training and optimization cost.

Model training was conducted using multi-GPU hardware to support cross-validation, extensive data augmentation, and reproducibility across multiple datasets. These training-time requirements represent a one-time offline cost and do not reflect the resources required for deployment. Similarly, PSO-based tuning operates in a low-dimensional parameter space and is executed for a limited number of iterations during offline optimization, without affecting inference-time complexity.

Inference-time efficiency.

At inference time, SOFAN-UNet performs a single forward pass without iterative optimization or auxiliary processing. The incorporation of hybrid attention mechanisms introduces a modest increase in parameter count relative to the baseline U-Net, but does not alter the asymptotic complexity of inference. Importantly, inference does not depend on cross-validation, data augmentation, or PSO execution and can be performed on a single GPU.

Deployment perspective and clinical implications.

Although experiments were conducted on high-performance hardware to ensure fair comparison and statistical robustness, the proposed model does not rely on specialized hardware features during inference. The consistently high AUC, Dice Similarity Coefficient, and boundary adherence metrics observed across multiple datasets indicate stable generalization behavior, which is critical for deployment across heterogeneous scanners and patient populations. These characteristics suggest that SOFAN-UNet is suitable for further evaluation in clinically realistic settings.

Future deployment considerations.

Extending validation to external clinical cohorts or prospective data may further strengthen confidence in real-world applicability. Additionally, broader profiling under diverse hardware constraints could facilitate deployment in resource-constrained environments. These directions represent natural extensions of the current framework and do not detract from the demonstrated effectiveness of the proposed approach.

6 Conclusion

This study introduced SOFAN-UNet, a hybrid deep learning framework tailored for precise brain tumor segmentation in multimodal MRI. Enhancing the U-Net backbone with a dual attention mechanism—Squeeze-and-Excitation (SE) and Self-Attention—the model captures both local channel dependencies and global spatial features. A lightweight Particle Swarm Optimization (PSO)-based short-tuning strategy was employed to optimize attention gate parameters (e.g. scale weights and filter sizes), dynamically improving focus on tumor regions while preserving spatial detail through efficient skip connections. Evaluated on the FBTS, BraTS 2021, and BraTS 2018 datasets, SOFAN-UNet achieved a peak Dice Similarity Coefficient (DSC) of 0.9685, Jaccard Index (JI) of 0.9389, and AUC exceeding 0.999 in nearly all classes and modalities. Additionally, a low boundary error was recorded, with Hausdorff Distance (HD) as low as 1.0 and Average Symmetric Surface Distance (ASSD) reduced to 0.0041 in T1CE and Pituitary cases.

The model consistently outperformed the baseline U-Net and several state-of-the-art methods across datasets. In the FBTS dataset, it achieved DSCs of 0.9304 (Meningioma), 0.8053 (Glioma), and 0.8591 (Pituitary), while in BraTS 2021, the T1CE modality achieved DSC = 0.9691 and MCC = 0.9687. Even in challenging LGG-T1 cases, where HD reached 10.0 due to weak tumor contrast, the model maintained high recall (0.9707) and F1-score (0.8999). Ablation experiments showed a 9.26 p.p. gain in DSC and 9.98 p.p. in JI over vanilla U-Net, without increasing parameter complexity (33.2M) or inference time ($\approx 1300\,$s). These results validate SOFAN-UNet's adaptability, stability, and seg-

mentation precision across structural and contrast-enhanced modalities. Future work will extend to tumor sub-region analysis, attention regularization, and real-time deployment on low-power platforms to further bridge clinical translation.

A Hyperparameter Selection and Justification

The selection of hyperparameters plays a crucial role in the performance and stability of deep learning models, especially in medical image segmentation tasks that involve small datasets and high structural complexity. The following justifications are provided for each training and optimization parameter used in the SOFAN-UNet framework, based on literature best practices, empirical evaluations, and architectural constraints.

Learning Rate (1×10^{-4}). A fixed learning rate of 1×10^{-4} was used with the Adam optimizer throughout the experiments. This choice is supported by numerous medical imaging studies that use U-Net and its variants for segmentation. Ronneberger et al. [29] used a similar learning rate in the original U-Net, and Isensee et al. [126] also showed stability around this scale in nnU-Net. During preliminary trials, we observed that higher rates (1×10^{-3}) led to an oscillating validation loss, while lower rates (1×10^{-5}) resulted in slower convergence without a significant improvement in accuracy.

Batch Size (32). A batch size of 32 was selected to balance GPU memory utilization and gradient stability. Larger batch sizes (e.g., 64) led to memory overflow during attention optimization with PSO, whereas smaller batches (e.g., 8 or 16) resulted in noisy gradients and slower convergence. Batch sizes of 16-32 have been commonly reported in segmentation models such as DeepLabV3+ and U-Net, providing empirical support for our selection.

Epochs (50). We selected 50 epochs as a fixed limit for baseline training and partial training during PSO iterations. This value was experimentally validated to allow convergence of the attention-enhanced U-Net without excessive overfitting. Early stopping was used during experimentation, and convergence was typically achieved by epoch 40, with marginal improvements afterward.

PSO Particle Count (5) and Iterations (10). The PSO parameters were chosen on the basis of a trade-off between optimization quality and computational cost. A particle count of 5 is within the typical range used in metaheuristic tuning of neural networks, where values between 5–20 are common. Limiting to 10 iterations enabled a short optimization cycle (as the "SO" in SOFAN suggests) while still allowing the swarm to explore the bounded attention parameter space. These settings proved sufficient to find stable optimal values ϕ_s^* and ϕ_f^* with consistent validation Dice improvement (detailed in Section 5.1).

PSO Inertia and Acceleration Coefficients ($w = 0.5$, $c_1 = c_2 = 1.5$). The standard PSO dynamics were followed using an inertia weight $w = 0.5$, and symmetric cognitive/social components $c_1 = c_2 = 1.5$. These values were derived from Clerc and Kennedy's constriction coefficient framework [127], which has been widely adopted in swarm-based neural network tuning. They balance exploration and exploitation, stabilizing particle updates without erratic jumps.

Parameter Bounds for ϕ_s and ϕ_f ($[0.5, 3.0]$). The scale factor ϕ_s determines the compression ratio in the SE block, while ϕ_f controls the filter expansion in the spatial attention branch. The lower bound (0.5) ensures that the attention modules retain minimal expressiveness, avoiding collapse. The upper bound (3.0) limits excessive channel expansion that can lead to overfitting or GPU memory constraints. This range was informed by architectural studies of attention modules [128, 129] and was empirically validated in our short-range tuning.

Fitness Function (Validation Dice Coefficient). We selected the Dice similarity coefficient as the fitness function during PSO because it is highly sensitive to class imbalance and is a standard metric for segmentation quality. The Dice coefficient is also differentiable and provides smooth feedback during tuning, whereas the binary accuracy or IoU may plateau early in optimization.

We additionally conducted a sensitivity analysis across multiple PSO swarm sizes and iteration budgets (Section 3.2.4), confirming that the selected configuration achieves stable convergence with minimal computational overhead.

Table 24. Summary of SOFAN-UNet parameter choices and justification

Parameter	Value	Justification
Learning rate	1×10^{-4}	Stable convergence in U-Net literature and validated empirically
Batch size	32	Best trade-off between convergence, noise, and GPU efficiency
Epochs	50	Ensures convergence without overfitting; matches observed learning curves
PSO particles / iterations	5 / 10	Efficient exploration with minimal cost; consistent with PSO literature
Inertia / acceleration (w, c_1, c_2)	0.5 / 1.5 / 1.5	Based on PSO constriction theory and CNN tuning practices
ϕ_s, ϕ_f bounds	$[0.5, 3.0]$	Avoids collapse or overfitting; derived from attention studies
Fitness function	Dice coefficient	Sensitive to imbalance and segmentation accuracy

B Extended Quantitative Results

This appendix provides the full tabulated results for each dataset across all training, validation, testing, and five-fold cross-validation phases. These extended results complement the summarized findings in Section 4.3.1 and are provided here to avoid redundancy while preserving the traceability, detail, and transparency of the evaluation.

B.1 FBTS Dataset (Figshare Brain Tumor Segmentation)

The evaluation of the FBTS dataset offers critical insights into how tumor morphology and imaging characteristics affect segmentation performance across the training, validation, testing, and cross-validation stages.

Table 25 reports the accuracy of the model, the Dice Similarity Coefficient (DSC), the Jaccard Index (JI) and the loss for three types of tumors. In particular, across all phases, Meningioma and Pituitary tumors exhibit high DSC and JI values, accompanied by minimal loss—indicative of their relatively homogeneous texture and spatial confinement in the T1CE sequences. In contrast, Glioma yields comparatively lower validation and testing metrics, especially in JI. This discrepancy is attributable to its diffuse and heterogeneous growth patterns, which frequently overlap with normal tissue and challenge delineation boundaries. The elevated loss and reduced overlap scores reflect such ambiguity.

To further assess the model's reliability and variance across patient subsets, a five-fold cross-validation (5F-CV) was conducted. The fold-wise performance reported in Table 26 shows a notable consistency in Meningioma and Pituitary classes, whereas Glioma presents a broader range across folds. Specifically, Fold 1 shows a decrease in DSC and JI for Glioma, consistent with known clinical heterogeneity and interpatient variability. These fluctuations underscore the importance of evaluating deep models not just on average scores but across stratified patient groups.

Table 25. Training, validation, and testing results on FBTS dataset (per class).

Class	Acc	Loss	DSC	JI
Training				
Meningioma	0.9990	0.0038	0.9162	0.8461
Glioma	0.9989	0.0025	0.9649	0.9322
Pituitary	0.9996	0.0009	0.9664	0.9351
Validation				
Meningioma	0.9990	0.0036	0.9234	0.8579
Glioma	0.9918	0.0503	0.7996	0.6735
Pituitary	0.9981	0.0107	0.8828	0.7934
Testing				
Meningioma	0.9991	0.0035	0.9304	0.8699
Glioma	0.9921	0.0493	0.8053	0.6770
Pituitary	0.9981	0.0120	0.8591	0.7539

The statistical distribution of the Dice and Jaccard scores across folds (Table 27) highlights a strong central tendency and a low dispersion for Pituitary, followed by Meningioma. Glioma shows greater variability, as reflected in a wider IQR and slightly lower Q1 values. These metrics align with clinical insights: compact, well-defined tumors enable better boundary estimation, whereas irregular shapes introduce segmentation uncertainty.

Table 26. Five-fold cross-validation results on FBTS dataset (fold-wise summary).

Class	Fold	Acc	Loss	DSC	JI
Meningioma	1	0.9978	0.0064	0.9185	0.8493
	2	0.9976	0.0071	0.9071	0.8302
	3	0.9982	0.0062	0.9325	0.8738
	4	0.9984	0.0042	0.9423	0.8911
	5	0.9983	0.0071	0.9445	0.8951
Glioma	1	0.9905	0.0767	0.7341	0.5638
	2	0.9954	0.0196	0.8844	0.7932
	3	0.9969	0.0131	0.9214	0.8476
	4	0.9982	0.0063	0.9580	0.9190
	5	0.9985	0.0060	0.9610	0.9222
Pituitary	1	0.9979	0.0100	0.8576	0.7515
	2	0.9985	0.0060	0.8948	0.8103
	3	0.9987	0.0049	0.9246	0.8601
	4	0.9993	0.0021	0.9543	0.9126
	5	0.9995	0.0020	0.9640	0.9321

Table 27. 5F-CV summary statistics on FBTS dataset.

Class	Mean	Std	Q1	Q3
DSC				
Meningioma	0.8383	0.1785	0.7968	0.9672
Glioma	0.8918	0.0836	0.8844	0.9580
Pituitary	0.9191	0.0391	0.8948	0.9543
JI				
Meningioma	0.7588	0.2332	0.6672	0.9268
Glioma	0.8151	0.1254	0.7935	0.9195
Pituitary	0.8530	0.0660	0.8103	0.9126

Overall, the results validate the segmentation model's ability to handle inter-class variation while preserving high overlap and boundary alignment. The Meningioma and Pituitary classes demonstrate excellent agreement between predicted masks and ground truth. At the same time, Glioma's complexity presents segmentation limitations that could benefit from further spatial attention refinement or data stratification strategies.

B.2 BraTS 2021 Dataset

The BraTS 2021 dataset evaluation provides a modality-specific perspective on whole-tumor segmentation. The complete experimental results—covering training, validation, testing, and five-fold cross-validation (5F-CV)—are detailed in Tables 28, 29, and 30.

Table 28. Training, validation, and testing results on BraTS 2021 dataset (per modality).

Modality	Acc	Loss	DSC	JI
Training				
FLAIR	0.9982	0.0042	0.9574	0.9183
T1	0.9970	0.0074	0.9282	0.8678
T2	0.9982	0.0043	0.9565	0.9167
T1CE	0.9983	0.0041	0.9593	0.9218
Validation				
FLAIR	0.9944	0.0229	0.9022	0.8232
T1	0.9887	0.0501	0.8141	0.6884
T2	0.9924	0.0390	0.8681	0.7697
T1CE	0.9892	0.0583	0.8181	0.6933
Testing				
FLAIR	0.9938	0.0240	0.8969	0.8160
T1	0.9892	0.0453	0.8254	0.7032
T2	0.9927	0.0382	0.8737	0.7769
T1CE	0.9903	0.0560	0.8254	0.7034

Table 29. Five-fold cross-validation results on BraTS 2021 (fold-wise summary).

Modality	Fold	Acc	Loss	DSC	JI
FLAIR	1	0.9932	0.0256	0.8774	0.7830
	2	0.9942	0.0232	0.9016	0.8211
	3	0.9973	0.0082	0.9459	0.8974
	4	0.9972	0.0082	0.9459	0.8977
	5	0.9978	0.0069	0.9618	0.9264
T1	1	0.9881	0.0706	0.7967	0.6642
	2	0.9871	0.0517	0.7521	0.6036
	3	0.9967	0.0125	0.9363	0.8803
	4	0.9935	0.0198	0.8799	0.7862
	5	0.9977	0.0084	0.9612	0.9255
T2	1	0.9845	0.0563	0.7736	0.6322
	2	0.9951	0.0245	0.9052	0.8273
	3	0.9973	0.0094	0.9493	0.9035
	4	0.9977	0.0090	0.9579	0.9192
	5	0.9984	0.0056	0.9723	0.9462
T1CE	1	0.9879	0.0836	0.7902	0.6540
	2	0.9936	0.0364	0.8761	0.7804
	3	0.9965	0.0137	0.9343	0.8769
	4	0.9970	0.0126	0.9439	0.8940
	5	0.9971	0.0123	0.9515	0.9076

Performance variation among modalities reflects their intrinsic imaging contrasts. FLAIR and T2, which enhance peritumoral and edema boundaries, consistently yield high segmentation scores. In contrast, T1 and T1CE modalities, which focus on structural integrity and enhancing tumor cores, reveal more variability, especially in the validation and test DSC. The higher validation loss for

T1CE, despite high training performance, suggests a contrast-specific domain shift, likely due to variations in enhancement intensity across scans.

Cross-validation consistency is best observed in FLAIR, which maintains DSC above 0.93 in the majority of the folds. The standard deviation analysis in Table 30 shows minimal fluctuation in these scores, indicating that this modality provides the most stable imaging foundation for whole tumor segmentation. Although T1CE performs competitively, its higher variance suggests sensitivity to enhancement timing and intensity artifacts, which may warrant adaptive normalization or attention reweighting.

Table 30. 5F-CV summary statistics on BraTS 2021 dataset.

Modality	Mean	Std	Q1	Q3
		DSC		
FLAIR	0.9265	0.0317	0.9016	0.9459
T1	0.8653	0.0800	0.7967	0.9363
T2	0.9117	0.0726	0.9052	0.9579
T1CE	0.8992	0.0606	0.8761	0.9439
		JI		
FLAIR	0.8651	0.0539	0.8211	0.8977
T1	0.7719	0.1229	0.6642	0.8803
T2	0.8457	0.1138	0.8273	0.9192
T1CE	0.8226	0.0954	0.7804	0.8940

B.3 BraTS 2018 Dataset (HGG and LGG)

This section presents extended performance metrics of SOFAN-UNet on the BraTS 2018 dataset, separated into High-Grade Glioma (HGG) and Low-Grade Glioma (LGG) cases. Tables 31, 32, and 33 report the full results across training, validation, testing, and five-fold cross-validation (5F-CV).

The HGG sequences demonstrate consistently high performance during all phases, particularly in T1 and FLAIR, which typically provide clear tumor delineation. LGG results show greater variation across modalities, especially in T2 and T1CE, which are more susceptible to indistinct boundaries. This variation contributes to performance gaps in both Dice and Jaccard metrics across LGG cases.

Table 31. Training, validation, and testing results on BraTS 2018 dataset (HGG and LGG).

Subtype-Modality	Acc	Loss	DSC	JI
		HGG		
FLAIR	0.9959	0.0117	0.8849	0.7943
T1	0.9982	0.0055	0.9435	0.8933
T2	0.9972	0.0079	0.9232	0.8579
T1CE	0.9972	0.0078	0.9227	0.8569
		LGG		
FLAIR	0.9986	0.0052	0.9416	0.8901
T1	0.9987	0.0047	0.9462	0.8981
T2	0.9949	0.0355	0.6764	0.5139
T1CE	0.9950	0.0256	0.7451	0.5961

Table 32. Five-fold cross-validation results on BraTS 2018 dataset (fold-wise summary).

Subtype-Modality	Fold	Acc	Loss	DSC	JI
			HGG		
T1CE	1	0.9936	0.0197	0.8877	0.7982
	2	0.9955	0.0158	0.9323	0.8737
	3	0.9961	0.0112	0.9384	0.8740
	4	0.9938	0.0164	0.8755	0.7793
	5	0.9871	0.0340	0.7872	0.6492
T2	1	0.9940	0.0249	0.9114	0.8222
	2	0.9937	0.0221	0.8813	0.7809
	3	0.9882	0.0552	0.7957	0.6805
	4	0.9872	0.0325	0.7688	0.6068
	5	0.9842	0.0424	0.7186	0.5429
T1	1	0.9961	0.0112	0.9384	0.8740
	2	0.9938	0.0164	0.8755	0.7793
	3	0.9871	0.0340	0.7872	0.6492
	4	0.9753	0.0644	0.5806	0.4124
	5	0.9732	0.0670	0.5481	0.3786
			LGG		
T1CE	1	0.9977	0.0073	0.9558	0.9153
	2	0.9975	0.0082	0.9580	0.9194
	3	0.9958	0.0175	0.9211	0.8538
	4	0.9917	0.0494	0.8608	0.7556
	5	0.9786	0.0676	0.7106	0.5511
T2	1	0.9974	0.0086	0.9550	0.9139
	2	0.9963	0.0128	0.9261	0.8624
	3	0.9944	0.0206	0.9200	0.8518
	4	0.9826	0.1146	0.7454	0.5942
	5	0.9862	0.0845	0.7339	0.5796

Table 33. 5F-CV summary statistics on BraTS 2018 dataset (HGG and LGG).

Modality	DSC Mean	DSC Std	JI Mean	JI Std
		HGG		
FLAIR	0.8143	0.1453	0.7092	0.1794
T1	0.7460	0.1562	0.6208	0.1988
T2	0.8152	0.0714	0.6958	0.1024
T1CE	0.8075	0.0992	0.6899	0.1385
		LGG		
FLAIR	0.6680	0.1946	0.5326	0.2138
T1	0.7612	0.1001	0.6248	0.1278
T2	0.8935	0.0751	0.8152	0.1128
T1CE	0.9329	0.0201	0.8750	0.0355

Overall, the results reveal distinct patterns: for HGG, T1 and T2 performed best in consistent segmentation due to well-defined structures. For LGG, the highest performance was achieved in the T1CE and T2 modalities, benefiting from slightly improved contrast profiles despite inherent segmentation challenges. The cross-validated consistency supports the robustness of the proposed framework on both low- and high-grade gliomas under multi-modality inputs.

Note: ROC visualizations are provided in D, and performance comparisons across datasets are discussed in Section 4.3.1.

C Model Performance Metrics Distribution Across Tumor Classes and Modalities

This appendix presents a comprehensive visualization of per-image performance distributions for the SOFAN-UNet segmentation model across all evaluated classes and imaging modalities. The assessment encompasses several key evaluation metrics, including Area Under the Curve (AUC), Dice Similarity Coefficient (DSC), Jaccard Index (JI), Precision, Recall, F1 Score, and Matthews Correlation Coefficient (MCC). These metrics collectively provide a multi-faceted view of the model's segmentation performance.

Each boxplot in the appendix illustrates the distribution of these performance metrics across individual test images. These visualizations are categorized by tumor type in the FBTS dataset—namely Meningioma, Glioma, and Pituitary tumors—and by imaging modality in the BraTS 2021 and BraTS 2018 datasets, including T1, T2, FLAIR, and T1CE scans. The analysis further distinguishes between tumor grades, covering both low-grade gliomas (LGG) and high-grade gliomas (HGG).

The boxplots enable in-depth analysis of consistency, variance, and potential outliers in segmentation quality. This graphical approach offers valuable insights into the robustness and generalization capability of the SOFAN-UNet model under diverse and heterogeneous imaging conditions. Figures 11 to 13 provide detailed depictions of the evaluated metrics for all datasets.

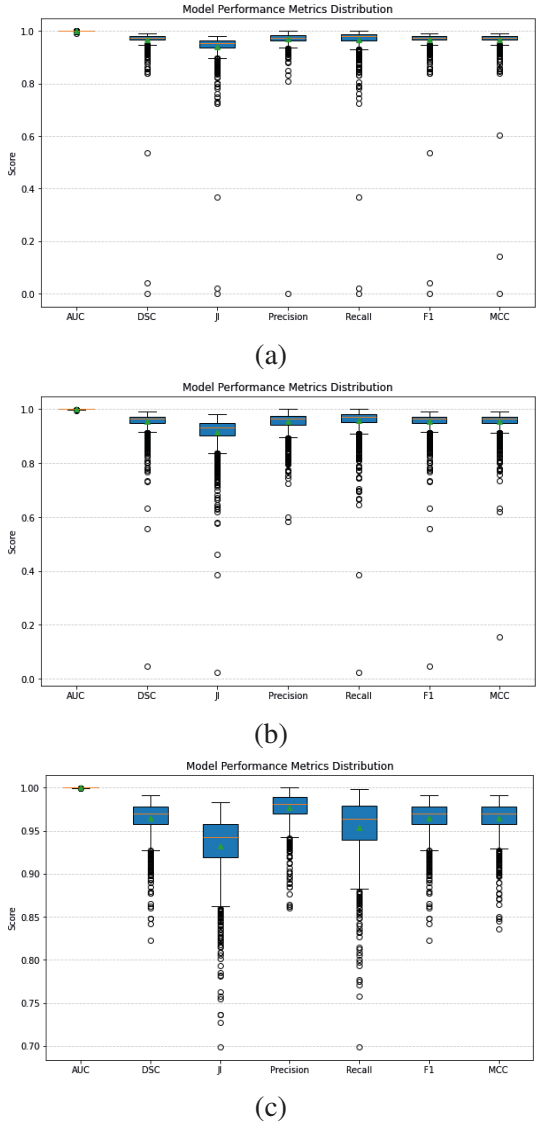


Figure 11. Performance metric distributions on the FBTS dataset: (a) Meningioma, (b) Glioma, and (c) Pituitary

Across all datasets and modalities, the AUC values remain consistently high (approaching 1.0), reflecting the model's strong discriminative ability to distinguish tumor from background voxels.

For the FBTS dataset, all three tumor types—Meningioma, Glioma, and Pituitary—demonstrated tightly clustered boxplots for DSC, JI, Precision, Recall, and F1, indicating stable performance with minimal variation. Slightly lower JI values in Glioma cases suggest occasional boundary ambiguity, as reflected in a slightly wider interquartile range in recall.

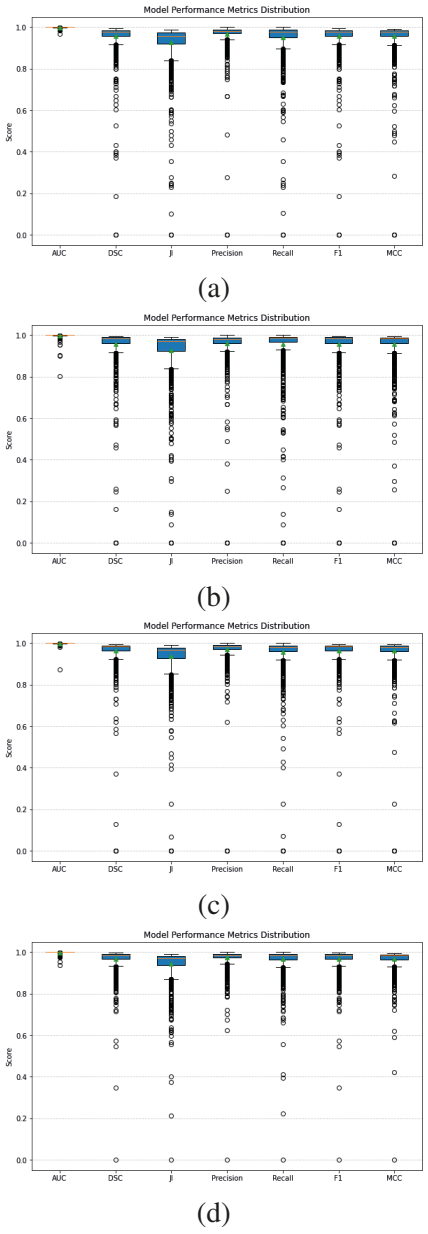


Figure 12. Performance metric distributions for each modality in BraTS 2021: (a) FLAIR, (b) T1, (c) T2, (d) T1CE

In the BraTS 2021 modalities, T1CE and T2 achieved the highest median scores across all metrics, especially DSC and Precision, indicating that contrast-enhanced sequences contribute to better tumor boundary delineation. FLAIR and T1 show slightly broader dispersion, particularly in IoU and Recall, which may be related to edema or non-enhancing tumor components that blend with surrounding tissue.

The BraTS 2018 HGG cases exhibit greater variance in JI and MCC than the LGG cases, particularly on FLAIR and T2, reflecting increased heterogeneity and necrosis, which are common in high-grade gliomas. The LGG subset shows more stable scores, especially in T1CE and T1, where the boxplots are compact with minimal outliers, suggesting cleaner boundaries and more consistent model behavior.

Overall, the model maintains high median scores across datasets, with robustness seen in consistent MCC and F1 distributions. Differences in performance across modalities and tumor grades highlight the importance of selecting imaging sequences, with contrast-enhanced and anatomical modalities generally yielding higher segmentation confidence.

These findings underscore the model's ability to generalize across various imaging contexts, especially when trained on anatomically rich or contrast-enhanced sequences.

D ROC-AUC Curves and Extended Analysis

This appendix presents visualizations of the ROC curves corresponding to each dataset and modality evaluated in this study. The analysis focuses on how SOFAN-UNet distinguishes tumor from background across varying thresholds, and how model confidence shifts in response to data variability. The AUC values reflect the threshold-independent classification quality and are interpreted through the lens of the modality sensitivity and tumor characteristics. These plots complement the AUC summaries presented in Section 4.3.

D.1 FBTS Dataset—Tumor Class Confidence Trends

Figure 14 illustrates the ROC curves for Meningioma, Glioma, and Pituitary tumors. The curves rise sharply toward the top-left corner, confirming the model's ability to separate positive and negative pixel distributions across tumor types.

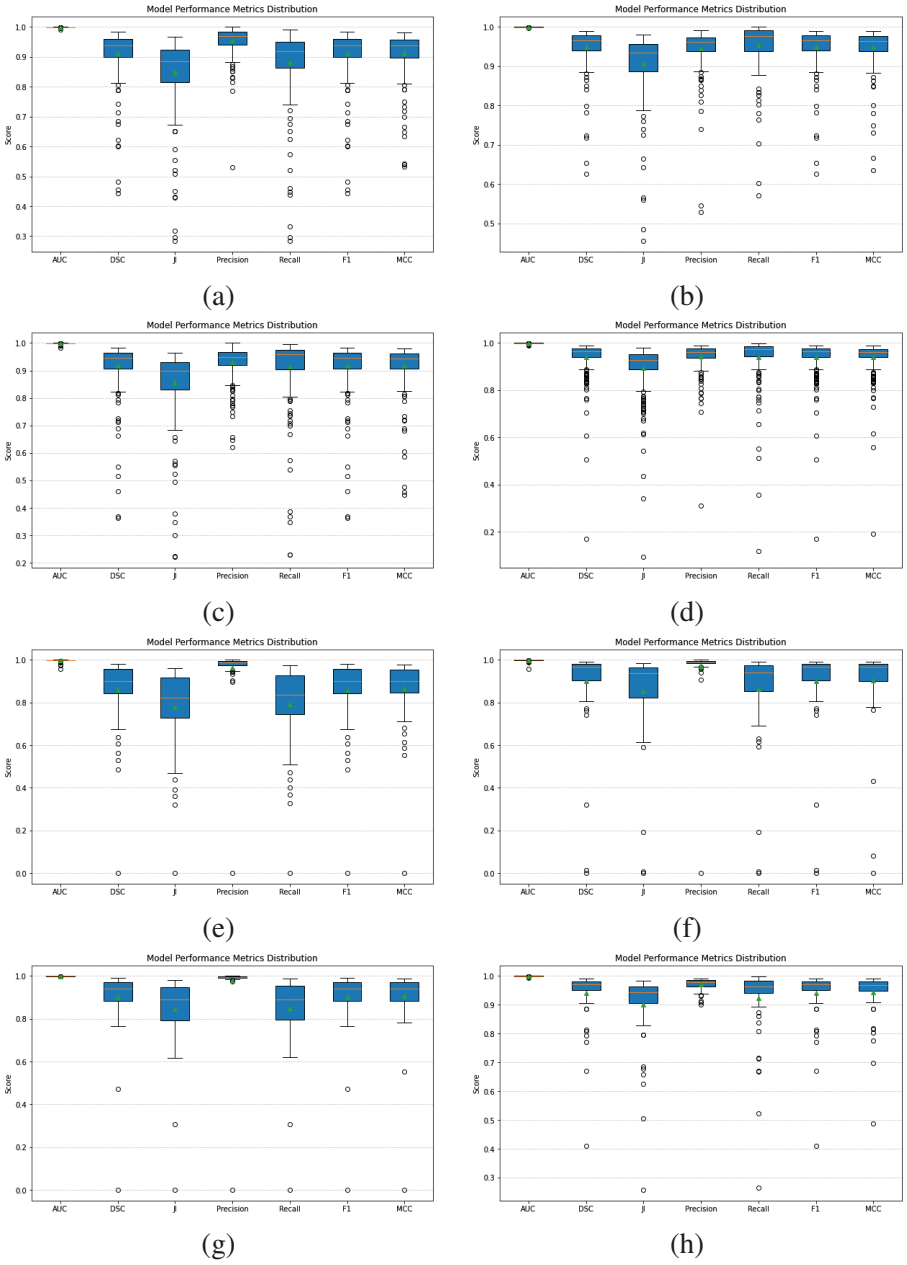


Figure 13. Performance metric distributions for BraTS 2018 (HGG and LGG): (a) FLAIR, (b) T1, (c) T2, (d) T1CE, respectively

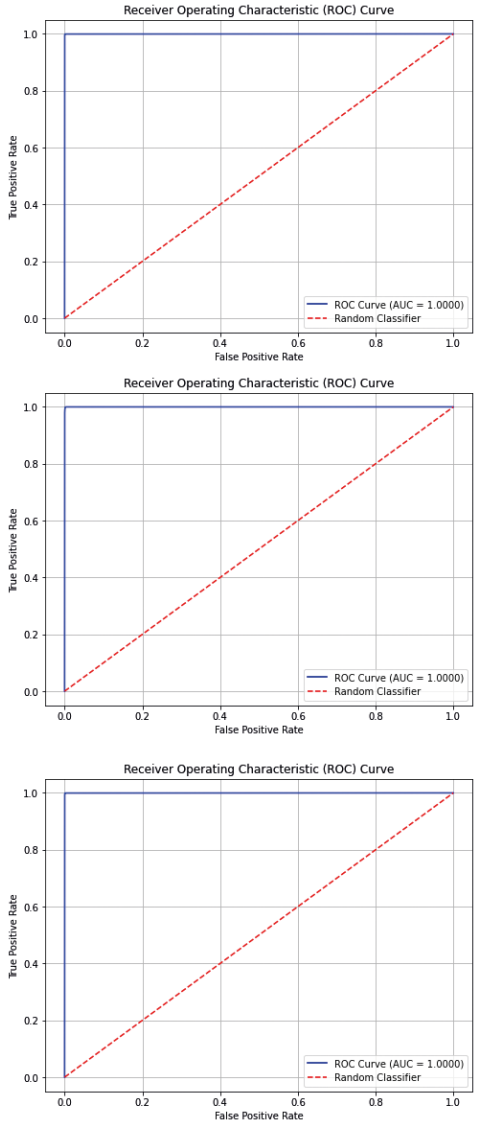


Figure 14. ROC curves for FBTS tumor classes: Meningioma (top-left), Glioma (top-right), and Pituitary (bottom).

For Meningioma and Pituitary, the curves approach a nearly perfect vertical ascent at low false positive rates. This profile indicates highly separable class distributions and well-calibrated probability outputs. The Glioma shows a slightly more gradual curvature, suggesting greater overlap between the predicted probabilities for the tumor and the background. This aligns with the inherent visual complexity of gliomas, which often display diffuse and infiltrative growth patterns that reduce the certainty of edge delineation.

D.2 BraTS 2021—Modality-Driven Classification Profiles

The ROC curves in Figure 15 correspond to the BraTS 2021 dataset segmented by MRI modality: FLAIR, T1, T1CE, and T2.

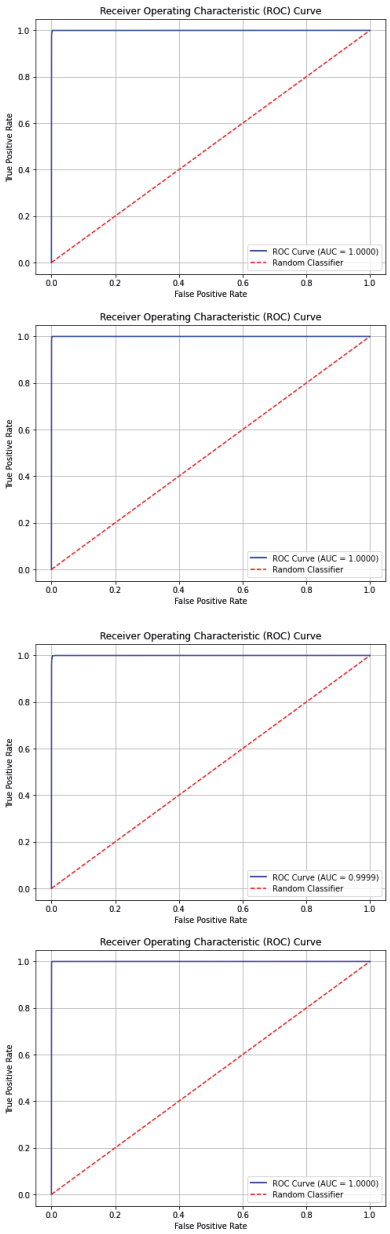


Figure 15. ROC curves for BraTS 2021 modalities: FLAIR, T1, T1CE, and T2.

FLAIR achieves the most vertical and top-heavy curve, revealing strong sensitivity across a wide range of thresholds. This is expected, as FLAIR is particularly effective in visualizing edema and diffuse tumor regions, which dominate the whole tumor label. T1CE also displays a robust curve shape, particularly in the mid- to

high-threshold regions, benefiting from contrast enhancement in active tumor zones.

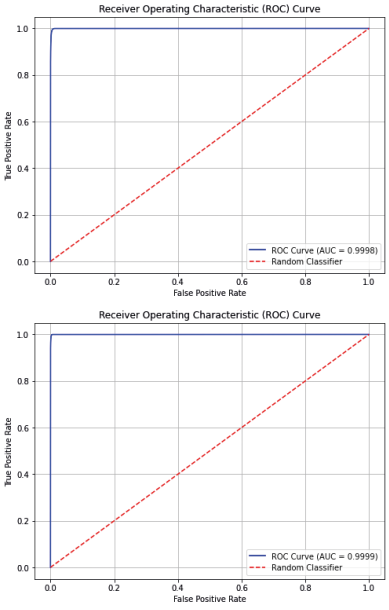
The T1 and T2 curves exhibit slightly shallower slopes during early threshold progression, implying greater uncertainty in classifying marginal tumor voxels. These patterns are consistent with lower tumor-to-background contrast in native T1 and fluid-sensitive T2 images.

D.3 BraTS 2018—Grade and Modality Influences on Class Boundaries

Figures 16 and 17 show the ROC curves for high-grade gliomas (HGG) and low-grade gliomas (LGG) from BraTS 2018.

In HGG cases, all modalities show tightly clustered ROC curves that rise sharply with minimal deviation. T1 and T1CE yield the steepest profiles, consistent with the strong tumor delineation provided by their intensity contrasts in enhancing tumor regions. FLAIR and T2 curves closely follow, reflecting their utility in detecting surrounding edema and infiltrative components.

In LGG, although AUC values remain high ($>$ 0.9995), there is a slightly broader curvature, particularly in T1CE and T2. These shapes suggest a reduction in predictive confidence at intermediate thresholds. This pattern corresponds to the typically less aggressive and more homogeneous tissue characteristics of LGG tumors, which reduce signal variation and edge distinction.



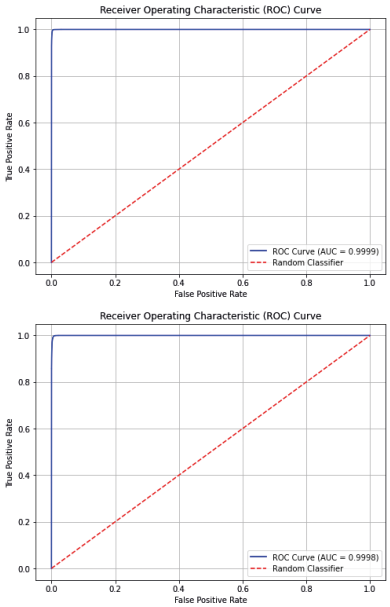
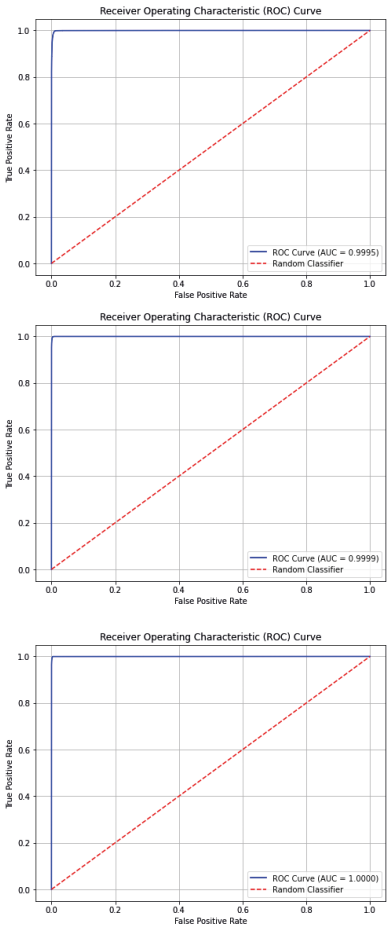


Figure 16. ROC curves for BraTS 2018 HGG cases: FLAIR, T1, T1CE, and T2.



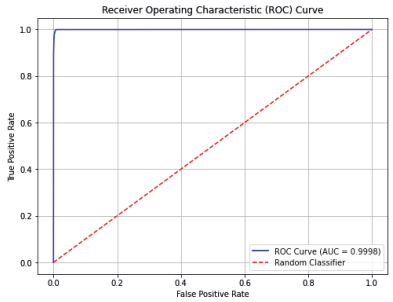


Figure 17. ROC curves for BraTS 2018 LGG cases: FLAIR, T1, T1CE, and T2.

Overall, these visualizations emphasize the model's ability to scale its confidence under different image contrast mechanisms and tumor morphologies. ROC profiles confirm that the PSO-tuned attention mechanism allows SOFAN-UNet to maintain consistent classification strength across subtle tumor variations and cross-modal scenarios.

D.4 Summary of ROC-AUC Value Ranges

For convenience, Table 34 provides a consolidated summary of all ROC-AUC intervals observed.

Table 34. Summary of ROC-AUC ranges across datasets and modalities (pixel-level classification)

Dataset / Class / Modality	Type	ROC-AUC Range
FBTS – Meningioma	Tumor Class	0.9994 – 1.0000
FBTS – Glioma	Tumor Class	0.9999 – 1.0000
FBTS – Pituitary	Tumor Class	1.0000 – 1.0000
BraTS 2021 – FLAIR	Modality	0.9997 – 1.0000
BraTS 2021 – T1	Modality	0.9999 – 1.0000
BraTS 2021 – T2	Modality	0.9999 – 1.0000
BraTS 2021 – T1CE	Modality	0.9999 – 1.0000
BraTS 2018 – FLAIR (HGG)	Modality	0.9994 – 1.0000
BraTS 2018 – T1 (HGG)	Modality	0.9993 – 1.0000
BraTS 2018 – T2 (HGG)	Modality	0.9989 – 1.0000
BraTS 2018 – T1CE (HGG)	Modality	0.9993 – 1.0000
BraTS 2018 – FLAIR (LGG)	Modality	0.9985 – 0.9999
BraTS 2018 – T1 (LGG)	Modality	0.9998 – 1.0000
BraTS 2018 – T2 (LGG)	Modality	0.9998 – 1.0000
BraTS 2018 – T1CE (LGG)	Modality	0.9999 – 1.0000

These ROC-AUC intervals and curve patterns complement Dice-based segmentation analysis by offering an additional confidence layer in evaluating model discriminability across tumor categories and imaging contrasts.

E Augmentation Evaluation Supplementary Analysis on FBTS Dataset

This appendix provides the full numerical results corresponding to the visual analysis in Section 4.2. The tabulated data below summarizes the absolute performance metrics of the segmentation models trained with and without augmentation, evaluated across training, validation, and testing phases for all tumor classes in the FBTS dataset.

The tabulated values complement the visual results shown in Figure 5 and Figure 6, offering a complete breakdown of performance improvements across all evaluation phases.

Evaluation of augmentation effects on segmentation performance across Meningioma, Glioma, and Pituitary classes reveals clear patterns in both spatial accuracy and generalization ability. For the Meningioma class, results remained stable across all phases. In training, performance was already near saturation, with accuracy greater than 0.9985 and only slight variation in the Dice and Jaccard indices after augmentation. During validation, moderate gains were observed, with Dice increasing from 0.9162 to 0.9234 and Jaccard from 0.8457 to 0.8579. In the testing phase, a marginal increase in segmentation quality was observed, indicating that the augmentation acted primarily as a stabilizing factor for an already well-learned class.

For Glioma, the impact of the augmentation was substantial. In the training phase, Dice rose from 0.8384 to 0.9649 and Jaccard from 0.7253 to 0.9322, indicating a significant improvement in spatial representation learning. These improvements were sustained during validation, with Dice reaching 0.7996 and Jaccard 0.6735, despite a slight increase in loss. In testing, the augmentation effect was the most prominent, elevating Dice from 0.6880 to 0.8053 and Jaccard from 0.5271 to 0.6770. This indicates a strengthened capacity of the model to handle the structural irregularities and spatial ambiguity typical of Glioma tumors.

The Pituitary class also benefited from augmentation, particularly in spatial metrics. Training Dice improved from 0.8799 to 0.9664 and Jaccard from 0.7862 to 0.9351. Validation metrics followed this trend, with Dice increasing to 0.8828 and Jaccard to

Table 35. Performance comparison of FBTS models with and without augmentation across training, validation, and testing.

Class	Metric	Training				Validation				Testing			
		Acc	Loss	DSC	JI	Acc	Loss	DSC	JI	Acc	Loss	DSC	JI
Meningioma	w/o Aug	0.9985	0.0035	0.9381	0.8836	0.9976	0.0077	0.9162	0.8457	0.9980	0.0054	0.9276	0.8653
	w/ Aug	0.9990	0.0038	0.9162	0.8461	0.9990	0.0036	0.9234	0.8579	0.9991	0.0035	0.9304	0.8699
Glioma	w/o Aug	0.9954	0.0117	0.8384	0.7253	0.9896	0.0336	0.7147	0.5608	0.9889	0.0381	0.6880	0.5271
	w/ Aug	0.9989	0.0025	0.9649	0.9322	0.9918	0.0503	0.7996	0.6735	0.9921	0.0493	0.8053	0.6770
Pituitary	w/o Aug	0.9995	0.0025	0.8799	0.7862	0.9980	0.0083	0.8199	0.6956	0.9981	0.0083	0.8043	0.6743
	w/ Aug	0.9996	0.0009	0.9664	0.9351	0.9981	0.0107	0.8828	0.7934	0.9981	0.0120	0.8591	0.7539

0.7934. In the testing phase, performance remained strong, achieving 0.8591 for Dice and 0.7539 for Jaccard. Although slight increases in loss were recorded, augmentation clearly contributed to more robust and consistent boundary delineation across diverse tumor shapes.

Overall, training-phase gains reflect the model's enhanced learning of morphological patterns, while validation and test improvements confirm improved generalization under real-world variability. Augmentation proved to be the most impactful for the Glioma and Pituitary classes, which exhibit greater spatial heterogeneity. Meningioma, with its relatively uniform morphology, exhibited steadier but still measurable improvement. These observations highlight the effectiveness of augmentation in improving model adaptability and segmentation precision across tumor types and evaluation phases.

Acknowledgement

Research funding was provided by AGH University of Krakow (Program "Excellence initiative – research university"), ACK Cyfronet AGH (Grant no. PLG/2024/017503 and PLG/2025/018784), Ahmad Dahlan University (grant number: U12/310/III/2025 and U12/1319/XII/2025), and Polish Ministry of Science and Higher Education funds assigned to AGH University of Krakow.

References

[1] S. Khalighi, K. Reddy, A. Midya, K. B. Pandav, A. Madabhushi, and M. Abedalthagafi, Artificial intelligence in neuro-oncology: advances and challenges in brain tumor diagnosis, prognosis, and precision treatment, npj Precision Oncology, vol. 8, no. 1, p. 80, 2024.

[2] S. K. Mathivanan, S. Sonaimuthu, S. Murugesan, H. Rajadurai, B. D. Shivahare, and M. A. Shah, Employing deep learning and transfer learning for accurate brain tumor detection, Scientific Reports, vol. 14, no. 1, p. 7232, 2024.

[3] A. Batool and Y.-C. Byun, Brain tumor detection with integrating traditional and computational intelligence approaches across diverse imaging modalities - Challenges and future directions, Computers in Biology and Medicine, p. 108412, 2024.

[4] Jyoti and A. Kumar, A systematic review of trending technologies in non-invasive automatic brain tumor detection, Multimedia Tools and Applications, 2024.

[5] T. Zhou, Boundary-aware and cross-modal fusion network for enhanced multi-modal brain tumor segmentation, Pattern Recognition, vol. 165, p. 111637, 2025.

[6] P. R, J. P. P. M, and N. J S, Brain tumor segmentation using multi-scale attention U-Net with EfficientNetB4 encoder for enhanced MRI analysis, Scientific Reports, vol. 15, no. 1, p. 9914, 2025.

[7] M. A. Abid and K. Munir, A systematic review on deep learning implementation in brain tumor segmentation, classification and prediction, Multimedia Tools and Applications, 2025.

[8] S. Cornelissen, S. M. Schouten, P. P. J. H. Langenhuizen, S. T. Lie, H. P. M. Kunst, P. H. N. de With, and J. B. Verheul, Defining tumor growth in vestibular schwannomas: a volumetric inter-observer variability study in contrast-enhanced T1-weighted MRI, Neuroradiology, vol. 66, no. 11, pp. 2033–2042, 2024.

[9] Q. D. Dinh, D. Kunk, T. Son Hy, V. Nalam, and P. D. Dao, Machine learning for automated electrical penetration graph analysis of aphid feeding behavior: Accelerating research on insect-plant interactions, PLOS ONE, vol. 20, no. 4, p. e0319484, 2025.

[10] W. Jiangtao, N. I. R. Ruhaiyem, and F. Panpan, A Comprehensive Review of U-Net and Its Variants: Advances and Applications in Medical Image Segmentation, IET Image Processing, vol. 19, no. 1, 2025.

[11] R. Azad, E. K. Aghdam, A. Rauland, Y. Jia, A. H. Avval, A. Bozorgpour, S. Karimijafarbigloo, J. P. Cohen, E. Adeli, and D. Merhof, Medical Image Segmentation Review: The Success of U-Net, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 46, no. 12, pp. 10076–10095, 2024.

[12] Y. Peng, D. Z. Chen, and M. Sonka, U-Net V2: Rethinking the Skip Connections of U-Net for Medical Image Segmentation, in 2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI), pp. 1–5, IEEE, 2025.

[13] J. Chen, J. Mei, X. Li, Y. Lu, Q. Yu, Q. Wei, X. Luo, Y. Xie, E. Adeli, Y. Wang, M. P. Lungren, S. Zhang, L. Xing, L. Lu, A. Yuille, and Y. Zhou, TransUNet: Rethinking the U-Net architecture design for medical image segmentation through the lens of transformers, Medical Image Analysis, vol. 97, p. 103280, 2024.

[14] M. N. Hasan, A. Ishraq, A. Alam Emon, J. Shin, and M. M. Kabir, Advancing Breast Cancer Diagnosis: Attention-Enhanced U-Net for Breast Cancer Segmentation, in Mridha, M.F., Dey, N. (eds) Data-Driven Clinical Decision-Making Using Deep Learning in Imaging. Studies in Big Data, vol. 152, pp. 207–226, Springer, Singapore, 2024.

[15] Y. Xu, R. Quan, W. Xu, Y. Huang, X. Chen, and F. Liu, Advances in Medical Image Segmentation: A Comprehensive Review of Traditional, Deep Learning and Hybrid Approaches, Bioengineering, vol. 11, no. 10, p. 1034, 2024.

[16] M. A. Talukder, M. A. Layek, M. A. Hossain, M. A. Islam, M. Nur-e Alam, and M. Kazi, ACU-Net: Attention-based convolutional U-Net model for segmenting brain tumors in fMRI images, DIGITAL HEALTH, vol. 11, 2025.

[17] K. Pani and I. Chawla, Synthetic MRI in action: A novel framework in data augmentation strategies for robust multi-modal brain tumor segmentation, Computers in Biology and Medicine, vol. 183, p. 109273, 2024.

[18] S. Saifullah, R. Dreżewski, A. Yudhana, M. Wielgosz, and W. Caesarendra, Modified U-Net with attention gate for enhanced automated brain tumor segmentation, Neural Computing and Applications, vol. 37, no. 7, pp. 5521–5558, 2025.

[19] M. B. Hossain, R. K. Shinde, S. Oh, K.-C. Kwon, and N. Kim, A Systematic Review and Identification of the Challenges of Deep Learning Techniques for Undersampled Magnetic Resonance Image Reconstruction, Sensors, vol. 24, no. 3, p. 753, 2024.

[20] R. Teng, Y. Liao, W. Wang, Q. Duan, J. Liu, and F. Jin, MambaSC: A Feature Matching Method Using Mamba2 with Self and Cross-Attention for Multimodal Images, Journal of Artificial Intelligence and Soft Computing Research, vol. 16, pp. 163–184, mar 2026.

[21] S. Saifullah and R. Dreżewski, Particle Swarm-Optimized U-Net Framework for Precise Multimodal Brain Tumor Segmentation, in GECCO '25 Companion, (Malaga, Spain), ACM Press, 2025.

[22] A. Kunjumon, C. Jacob, and R. Resmi, Three-Dimensional Network With Squeeze and Excitation for Accurate Multi-Region Brain Tumor Segmentation, International Journal of Imaging Systems and Technology, vol. 35, no. 2, 2025.

[23] P. K. Dash and D. S. Sisodia, SEIMB-NET: A Squeeze and Excitation Driven Lightweight Model for Classification of Brain Tumors Using Magnetic Resonance Imaging, in 2024 2nd World Conference on Communication & Computing (WCONF), pp. 1–6, IEEE, 2024.

[24] F. Yang and B. Wang, Dual Channel-Spatial Self-Attention Transformer and CNN synergy network for 3D medical image segmentation, Applied Soft Computing, vol. 167, p. 112255, 2024.

[25] R. Azad, L. Niggemeier, M. Huttemann, A. Kazerouni, E. K. Aghdam, Y. Velichko, U. Bagci, and D. Merhof, Beyond Self-Attention: Deformable Large Kernel Attention for Medical Image Segmentation, in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 1287–1297, 2024.

[26] Y. Zhou, K. Ma, Q. Sun, Z. Wang, and M. Liu, Edge-Guided Cell Segmentation on Small Datasets Using an Attention-Enhanced U-Net Architecture, Information, vol. 15, no. 4, p. 198, 2024.

[27] Z. Zhu, Z. Wang, G. Qi, N. Mazur, P. Yang, and Y. Liu, Brain tumor segmentation in MRI with multi-modality spatial information enhancement and boundary shape correction, Pattern Recognition, vol. 153, p. 110553, 2024.

[28] M. S. S, V. G, S. N, S. S, R. R. R, and M. Yousuf, Enhanced Brain Tumors Segmentation In 3D Multimodal Images Using U-Net3+ And SegNet, in 2025 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI), pp. 1–6, IEEE, 2025.

[29] O. Ronneberger, P. Fischer, and T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, in Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) Medical Image Computing and

Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science(), pp. 234–241, Springer, Cham, 2015.

[30] Z. Fanyang and Z. Fan, CS U-NET : A Medical Image Segmentation Method Integrating Spatial and Contextual Attention Mechanisms Based on U-NET, International Journal of Imaging Systems and Technology, vol. 35, no. 2, 2025.

[31] H. Wang, P. Cao, J. Yang, and O. Zaiane, Narrowing the semantic gaps in U-Net with learnable skip connections: The case of medical image segmentation, Neural Networks, vol. 178, p. 106546, 2024.

[32] S. Dhivya, S. Mohanavalli, K. B. Sundharakumar, and I. Thamarai, Attention U^2Net: Cascaded UNets with Modified Skip Connection for Breast Tumor Segmentation, Neural Processing Letters, vol. 55, no. 9, pp. 11863–11883, 2023.

[33] N. S. Punn and S. Agarwal, Modality specific U-Net variants for biomedical image segmentation: a survey, Artificial Intelligence Review, vol. 55, no. 7, pp. 5845–5889, 2022.

[34] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications, IEEE Access, vol. 9, pp. 82031–82057, 2021.

[35] H. Abdel-Nabi, M. Ali, A. Awajan, M. Daoud, R. Alazrai, P. N. Suganthan, and T. Ali, A comprehensive review of the deep learning-based tumor analysis approaches in histopathological images: segmentation, classification and multi-learning tasks, Cluster Computing, vol. 26, no. 5, pp. 3145–3185, 2023.

[36] Y. Yu, S. Chen, and H. Wei, Modified UNet with attention gate and dense skip connection for flow field information prediction with porous media, Flow Measurement and Instrumentation, vol. 89, p. 102300, 2023.

[37] Q. Zhang, Y. Hang, J. Qiu, and H. Chen, Application of U-Net Network Utilizing Multiattention Gate for MRI Segmentation of Brain Tumors, Journal of Computer Assisted Tomography, vol. 48, no. 6, pp. 991–997, 2024.

[38] Y.-C. Liu, M. Shahid, W. Sarapugdi, Y.-X. Lin, J.-C. Chen, and K.-L. Hua, Cascaded atrous dual attention U-Net for tumor segmentation, Multimedia Tools and Applications, vol. 80, no. 20, pp. 30007–30031, 2021.

[39] S. P. Shyam, R. K. Bhuiyan, S. S. Mim, and C. M. A. Rahman, An Advanced Cascaded UNet Architecture with Dual Attention for Bio-Medical Image Segmentation, in 2024 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE), pp. 001–006, IEEE, 2024.

[40] J. Roth, J. Keller, S. Franke, T. Neumuth, and D. Schneider, Multi-plane UNet++ Ensemble for Glioblastoma Segmentation, in Crimi, A., Bakas, S. (eds) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2021. Lecture Notes in Computer Science, vol. 12962, pp. 285–294, Springer, Cham, 2022.

[41] A. Hou, L. Wu, H. Sun, Q. Yang, H. Ji, B. Cui, and P. Ji, Brain Segmentation Based on UNet++ with Weighted Parameters and Convolutional Neural Network, in 2021 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA), pp. 644–648, IEEE, 2021.

[42] Y. Lyu and X. Tian, MWG-UNet++: Hybrid Transformer U-Net Model for Brain Tumor Segmentation in MRI Scans, Bioengineering, vol. 12, no. 2, p. 140, 2025.

[43] D. Maji, P. Sigedar, and M. Singh, Attention Res-UNet with Guided Decoder for semantic segmentation of brain tumors, Biomedical Signal Processing and Control, vol. 71, p. 103077, 2022.

[44] T. Cao, G. Wang, L. Ren, Y. Li, and H. Wang, Brain tumor magnetic resonance image segmentation by a multiscale contextual attention module combined with a deep residual UNet (MCA-ResUNet), Physics in Medicine & Biology, vol. 67, no. 9, p. 095007, 2022.

[45] S. Jiang, X. Chen, and C. Yi, SSA-UNet: Whole brain segmentation by U-Net with squeeze-and-excitation block and self-attention block from the 2.5D slice image, IET Image Processing, vol. 18, no. 6, pp. 1598–1612, 2024.

[46] L. Xiong, C. Yi, Q. Xiong, and S. Jiang, SEA-NET: medical image segmentation network based on spiral squeeze-and-excitation and attention modules, BMC Medical Imaging, vol. 24, no. 1, p. 17, 2024.

[47] S. Roy Choi, J. Lee, and M. Lee, OrgUNETR: Utilizing Organ Information and Squeeze and Excitation Block for Improved Tumor Segmentation, IEEE Access, vol. 12, pp. 84122–84133, 2024.

[48] Y. Zhao, Y. Jiang, Z. Qian, L. Nie, and K. Xia, Hippocampus Segmentation with Transformer-UNet Fusion and Squeeze-And-Excitation Block, in Huang, DS., Pan, Y., Zhang, Q. (eds) Advanced Intelligent Computing in Bioinformatics. ICIC 2024. Lecture Notes in Computer Science, vol. 14882, pp. 420–430, Springer, Singapore, 2024.

[49] Z. Zhu, M. Sun, G. Qi, Y. Li, X. Gao, and Y. Liu, Sparse Dynamic Volume TransUNet with multi-level edge fusion for brain tumor segmentation, Computers in Biology and Medicine, vol. 172, p. 108284, 2024.

[50] T. Deepa and C. D. V. S. Rao, TransU-Net: Glioma Sub-region Segmentation, in 2024 1st International Conference on Cognitive, Green and Ubiquitous Computing (IC-CGU), pp. 01–06, IEEE, 2024.

[51] X. Ni and J. Mo, WDE-TransUNet: Brain Tumor Segmentation in MRI Images Utilizing Detail Enhancement and Transformers, in Proceedings of the 2024 5th International Symposium on Artificial Intelligence for Medicine Science, (New York, NY, USA), pp. 90–95, ACM, 2024.

[52] A. Vinisha and R. Boda, Efficient brain tumour detection system by Cascaded Fully Convolutional Improved DenseNet with Attention-based Adaptive Swin Unet-derived segmentation strategy, Multimedia Systems, vol. 30, no. 3, p. 134, 2024.

[53] J. Chen, X. Zhang, R. Li, and P. Zhou, Swin-HAUnet: A Swin-Hierarchical Attention Unet For Enhanced Medical Image Segmentation, in Lin, Z., et al. Pattern Recognition and Computer Vision. PRCV 2024. Lecture Notes in Computer Science, vol. 15044, pp. 371–385, Springer, Singapore, 2025.

[54] A. K. Upadhyay and A. K. Bhandari, MaS-TransUNet: A Multiattention Swin Transformer U-Net for Medical Image Segmentation, IEEE Transactions on Radiation and Plasma Medical Sciences, vol. 9, no. 5, pp. 613–626, 2025.

[55] S. Saifullah, R. Dreżewski, A. Yudhana, W. Caesarendra, and N. Huda, Bio-Inspired Metaheuristics in Deep Learning for Brain Tumor Segmentation: A Decade of Advances and Future Directions, Information, vol. 16, no. 6, p. 456, 2025.

[56] M. Arif, A. Jims, A. F., O. Geman, M.-D. Craciun, and F. Leuciuc, Application of Genetic Algorithm and U-Net in Brain Tumor Segmentation and Classification: A Deep Learning Approach, Computational Intelligence and Neuroscience, vol. 2022, pp. 1–11, 2022.

[57] A. Kaur, L. Kaur, and A. Singh, GA-UNet: UNet-based framework for segmentation of 2D and 3D medical images applicable on heterogeneous datasets, Neural Computing and Applications, vol. 33, no. 21, pp. 14991–15025, 2021.

[58] S. Saifullah and R. Dreżewski, Automatic Brain Tumor Segmentation Using Convolutional Neural Networks: U-Net Framework with PSO-Tuned Hyperparameters, in Affenzeller, M., et al. Parallel Problem Solving from Nature – PPSN XVIII. PPSN 2024. Lecture Notes in Computer Science, vol 15150. Springer, Cham., pp. 333–351, Springer, Cham, 2024.

[59] A. Malik and G. G. Devarajan, Integrated Brain Tumor Detection: PSO-Guided Segmentation with U-Net and CNN Classification, Procedia Computer Science, vol. 235, pp. 3447–3457, 2024.

[60] S. Saifullah and R. Dreżewski, GA-UNet: Genetic Algorithm-Optimized Lightweight U-Net Architecture for Multi-Sequence Brain Tumor MRI Segmentation, IEEE Access, vol. 13, pp. 175010–175024, 2025.

[61] S. Saifullah and R. Dreżewski, Optimizing U-Net Architecture Using Differential Evolution for Brain Tumor Segmentation, in Computational Science – ICCS 2025. ICCS 2025. Lecture Notes in Computer Science (M. Lees and Others, eds.), vol. 15906, pp. 403–411, Cham: Springer, 2025.

[62] N. Rasool and J. I. Bhat, A Critical Review on Segmentation of Glioma Brain Tumor and Prediction of Overall Survival, Archives of Computational Methods in Engineering, 2024.

[63] J. Cheng, W. Huang, S. Cao, R. Yang, W. Yang, Z. Yun, Z. Wang, and Q. Feng, Enhanced Performance of Brain Tumor Classification via Tumor Region Augmentation and Partition, PLOS ONE, vol. 10, no. 10, p. e0140381, 2015.

[64] U. Baid, S. Ghodasara, S. Mohan, M. Bilello, E. Calabrese, E. Colak, K. Farahani, J. Kalpathy-Cramer, F. C. Kitamura, S. Pati, L. Prevedello, J. Rudie, C. Sako, R. Shinohara, T. Bergquist, R. Chai, J. Eddy, J. Elliott, W. Reade, T. Schaffter, T. Yu, J. Zheng, C. Davatzikos, J. Mongan, C. Hess, S. Cha, J. Villanueva-Meyer, J. B. Freymann, J. S. Kirby, B. Wiestler, P. Crivellaro, R. R. Colen, A. Kotrotsou, D. Marcus, M. Milchenko, A. Nazeri, H. Fathallah-Shaykh, R. Wiest, A. Jakab, M.-A. Weber, A. Mahajan, B. Menze, A. E. Flanders, and S. Bakas, RSNA-ASNR-MICCAI-BraTS-2021 Dataset, 2023.

[65] B. H. Menze, A. Jakab, S. Bauer, Kalpathy-Cramer, et al., The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS), IEEE Transactions on Medical Imaging, vol. 34, no. 10, pp. 1993–2024, 2015.

[66] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. S. Kirby, J. B. Freymann, K. Farahani, and C. Davatzikos, Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features, Scientific Data, vol. 4, no. 1, p. 170117, 2017.

[67] S. Bakas, M. Reyes, A. Jakab, S. Bauer, et al., Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge, 2018.

[68] S. Saifullah, R. Dreżewski, A. Yudhana, and A. P. Suryotomo, Automatic Brain Tumor Segmentation: Advancing U-Net With ResNet50 Encoder for Precise Medical Image Analysis, IEEE Access, vol. 13, pp. 43473–43489, 2025.

[69] S. Saifullah, R. Dreżewski, and A. Yudhana, Advanced brain tumor segmentation using DeepLabV3Plus with Xception encoder on a multi-class MR image dataset, Multimedia Tools and Applications, 2025.

[70] J. Nalepa, M. Marcinkiewicz, and M. Kawulok, Data Augmentation for Brain-Tumor Segmentation: A Review, Frontiers in Computational Neuroscience, vol. 13, 2019.

[71] S. Saifullah, B. Yuwono, H. C. Rustamaji, B. Saputra, F. A. Dwiyanto, and R. Dreżewski, Detection of Chest X-ray Abnormalities Using CNN Based on Hyperparameter Optimization, Engineering Proceedings, vol. 56, no. 1, p. 223, 2023.

[72] E. Goceri, Medical image data augmentation: techniques, comparisons and interpretations, Artificial Intelligence Review, vol. 56, no. 11, pp. 12561–12605, 2023.

[73] S. Muksimova, S. Umirzakova, N. Iskhakova, A. Khaitov, and Y. I. Cho, Advanced convolutional neural network with attention mechanism for Alzheimer's disease classification using MRI, Computers in Biology and Medicine, vol. 190, p. 110095, 2025.

[74] N. Rasool and J. I. Bhat, Brain tumour detection using machine and deep learning: a systematic review, Multimedia Tools and Applications, vol. 84, no. 13, pp. 11551–11604, 2024.

[75] M. G. Abbas Malik, A. Saeed, K. Shehzad, and M. Iqbal, DEF-SwinE2NET: Dual enhanced features guided with multi-model fusion for brain tumor classification using preprocessing optimization, Biomedical Signal Processing and Control, vol. 100, p. 107079, 2025.

[76] P. Kumar Tiwary, P. Johri, A. Katiyar, and M. K. Chhipa, Deep Learning-Based MRI Brain Tumor Segmentation With EfficientNet-Enhanced UNet, IEEE Access, vol. 13, pp. 54920–54937, 2025.

[77] K. Bhagyalaxmi and B. Dwarakanath, CDCG-UNet: Chaotic Optimization Assisted Brain Tumor Segmentation Based on Dilated Channel Gate Attention U-Net Model, Neuroinformatics, vol. 23, no. 2, p. 12, 2025.

[78] S. Saifullah and R. Dreżewski, Brain Tumor Segmentation Using Ensemble CNN-Transfer Learning Models: DeepLabV3plus and ResNet50 Approach, in In: Franco, L., de Mulatier, C., Paszynski, M., Krzhizhanovskaya, V.V., Dongarra, J.J., Sloot, P.M.A. (eds) Computational Science – ICCS 2024. ICCS 2024. Lecture Notes in Computer Science, vol 14835, pp. 340–354, Springer, Cham, 2024.

[79] X. Zhang, Y. Zhang, Y. He, and M. Li, S3diff: Semantic Fusion and Structure-Guided Global Generation from a Single Image with Diffusion Models, Journal of Artificial Intelligence and Soft Computing Research, vol. 16, pp. 39–53, jan 2026.

[80] M. Hayat, Squeeze & Excitation joint with Combined Channel and Spatial Attention for Pathology Image Super-Resolution, Franklin Open, vol. 8, p. 100170, 2024.

[81] S. Zhu, Y. Yan, L. Wei, Y. Li, T. Mao, X. Dai, and R. Du, SECA-Net: Squeezed-and-excitated contextual attention network for medical image segmentation, Biomedical Signal Processing and Control, vol. 97, p. 106704, 2024.

[82] S. Yan, B. Yang, and A. Chen, A differential network with multiple gated reverse attention for medical image segmentation, Scientific Reports, vol. 14, no. 1, p. 20274, 2024.

[83] M. Fiaz, M. Noman, H. Cholakkal, R. M. Anwer, J. Hanna, and F. S. Khan, Guided-attention and gated-aggregation network for medical image segmentation, Pattern Recognition, vol. 156, p. 110812, 2024.

[84] M. M. Rahman, M. Munir, and R. Marculescu, Emcad: Efficient multi-scale convolutional attention decoding for medical image segmentation, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11769–11779, June 2024.

[85] L. Abualigah, A. Sheikhan, A. M. Ikotun, R. A. Zitar, A. R. Alsoud, I. Al-Shourbaji, A. G. Hussien, and H. Jia, Particle swarm optimization algorithm: review and applications, in Metaheuristic Optimization Algorithms, pp. 1–14, Elsevier, 2024.

[86] L. Abualigah, Particle Swarm Optimization: Advances, Applications, and Experimental Insights, Computers, Materials & Continua, vol. 82, no. 2, pp. 1539–1592, 2025.

[87] J. Nayak, H. Swapnarekha, B. Naik, G. Dhiman, and S. Vimal, 25 Years of Particle Swarm Optimization: Flourishing Voyage of Two Decades, Archives of Computational Methods in Engineering, vol. 30, no. 3, pp. 1663–1725, 2023.

[88] H. N. Fakhouri, A. Hudaib, and A. Sleit, Multi-vector particle swarm optimization algorithm, Soft Computing, vol. 24, no. 15, pp. 11695–11713, 2020.

[89] S. Saifullah and R. Drezewski, Improved Brain Tumor Segmentation Using Modified U-Net based on Particle Swarm Optimization Image Enhancement, in Genetic and Evolutionary Computation Conference (GECCO '24 Companion), (Melbourne, VIC, Australia), pp. 611–614, ACM, New York, NY, USA, 2024.

[90] Y. Chen, H. Xu, J. Liu, M. Hou, Y. Li, S. Qiu, M. Sun, H. Zhao, and W. Deng, A Hybridizing-Enhanced Quantum-Inspired Differential Evolution Algorithm with Multi-Strategy for Complicated Optimization, Journal of Artificial Intelligence and Soft Computing Research, vol. 16, pp. 5–37, jan 2026.

[91] J. Bilski, B. Kowalczyk, L. Dymova, and M. Xiao, Accelerating Neural Network Training with FS-GQR: A Scalable and High-Performance Alternative to Adam, Journal of Artificial Intelligence and Soft Computing Research, vol. 15, pp. 95–113, mar 2025.

[92] Z. Qiao, A. A. Heidari, X. Zhao, and H. Chen, An Evolutionary Neural Architecture Search Method Accelerated by Multi-Fidelity Evaluation and Genetic Decision Controller, Journal of Artificial Intelligence and Soft Computing Research, vol. 15, pp. 413–446, oct 2025.

[93] S. Saifullah and R. Drężewski, Bayesian Optimization-Driven U-Net Architecture Tuning for Brain Tumor Segmentation, Engineering Proceedings, vol. 124, p. 22, feb 2026.

[94] D. Cheng, X. Gao, Y. Mao, B. Xiao, P. You, J. Gai, M. Zhu, J. Kang, F. Zhao, and N. Mao, Brain tumor feature extraction and edge enhancement algorithm based on U-Net network, Heliyon, vol. 9, no. 11, p. e22536, 2023.

[95] Z. Sobhaninia, S. Rezaei, N. Karimi, A. Emami, and S. Samavi, Brain Tumor Segmentation by Cascaded Deep Neural Networks Using Multiple Image Scales, 2020 28th Iranian Conference on Electrical Engineering (ICEE), pp. 1–4, 2020.

[96] P. Chatterjee, K. Das Sharma, and A. Chakrabarti, Weakly supervised learning in domain transfer scenario for brain lesion segmentation in MRI, Multimedia Tools and Applications, 2024.

[97] A. Ingle, M. Roja, M. Sankhe, and D. Patkar, Efficient segmentation and classification of the tumor using improved encoder-decoder architecture in brain MRI images, International journal of electrical and computer engineering systems, vol. 13, no. 8, pp. 643–651, 2022.

[98] S. Mayala et al., Brain Tumor Segmentation Based on Minimum Spanning Tree, Frontiers in Signal Processing, vol. 2, 2022.

[99] S. Roshan, J. Tanha, M. Zarrin, A. F. Babaei, H. Nikkhah, and Z. Jafari, A deep ensemble medical image segmentation with novel sampling method and loss function, Computers in Biology and Medicine, p. 108305, 2024.

[100] S. K. V. Rao and B. Lingappa, Image Analysis for MRI Based Brain Tumour Detection Using Hybrid Segmentation and Deep Learning Classification Technique, International Journal of Intelligent Engineering and Systems, vol. 12, no. 5, pp. 53–62, 2019.

[101] S. Saifullah and R. Drężewski, Redefining brain tumor segmentation: a cutting-edge convolutional neural networks-transfer learning approach, International Journal of Electrical and Computer Engineering (IJECE), vol. 14, no. 3, p. 2583, 2024.

[102] A. Murmu and P. Kumar, A novel Gateaux derivatives with efficient DCNN-Resunet method for segmenting multi-class brain tumor, Medical & Biological Engineering & Computing, vol. 61, no. 8, pp. 2115–2138, 2023.

[103] S. Tripathi, A. Verma, and N. Sharma, Automatic segmentation of brain tumour in MR images using an enhanced deep learning approach, Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, vol. 9, no. 2, pp. 121–130, 2021.

[104] S. Kumar Bhatt, D. S. Srinivasan, and P. Prakash, Brain Tumor Segmentation Pipeline Model Using U-Net Based Foundation Model, Data and Metadata, vol. 2, p. 197, 2023.

[105] R. Mojtahedi, M. Hamghalam, and A. L. Simpson, Multi-modal Brain Tumour Segmentation Using Transformer with Optimal Patch Size, Bakas, S., et al. Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2022. Lecture Notes in Computer Science, vol. 13769, pp. 195–204, 2023.

[106] Z. Jiang, C. Ding, M. Liu, and D. Tao, Two-Stage Cascaded U-Net: 1st Place Solution to BraTS Challenge 2019 Segmentation Task, in Crimi, A., Bakas, S. (eds) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2019. Lecture Notes in Computer Science, pp. 231–241, Springer, Cham, 2020.

[107] F. D. Hernandez-Gutierrez, E. G. Avina-Bravo, D. F. Zambrano-Gutierrez, O. Almanza-Conejo, M. A. Ibarra-Manzano, J. Ruiz-Pinales, E. Ovalle-Magallanes, and J. G. Avina-Cervantes, Brain Tumor Segmentation from Optimal MRI Slices Using a Lightweight U-Net, Technologies, vol. 12, no. 10, p. 183, 2024.

[108] P. Bouchet et al., An Efficient Cascade of U-Net-Like Convolutional Neural Networks Devoted to Brain Tumor Segmentation, in Bakas, S., et al. Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2022. Lecture Notes in Computer Science, vol. 13769, pp. 149–161, Springer, Cham, 2023.

[109] M. S. Sadique, M. M. Rahman, W. Farzana, A. Temtam, and K. M. Iftekharuddin, Brain Tumor Segmentation Using Neural Ordinary Differential Equations with UNet-Context Encoding Network, in Bakas, S., et al. Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2022. Lecture Notes in Computer Science, pp. 205–215, Springer, Cham, 2023.

[110] S. Vijay, T. Guhan, K. Srinivasan, P. M. D. R. Vincent, and C.-Y. Chang, MRI brain tumor segmentation using residual Spatial Pyramid Pooling-powered 3D U-Net, Frontiers in Public Health, vol. 11, 2023.

[111] F. Ghazouani, P. Vera, and S. Ruan, Efficient brain tumor segmentation using Swin transformer and enhanced local self-attention, International Journal of Computer Assisted Radiology and Surgery, vol. 19, no. 2, pp. 273–281, 2023.

[112] J. Sachdeva, D. Sharma, and C. K. Ahuja, Multiscale segmentation net for segregating heterogeneous brain tumors: Gliomas on multimodal MR images, Image and Vision Computing, vol. 149, p. 105191, 2024.

[113] R. Yousef, S. Khan, G. Gupta, B. M. Albahlal, S. A. Alajlan, and A. Ali, Bridged-U-Net-ASPP-EVO and Deep Learning Optimization for Brain Tumor Segmentation, Diagnostics, vol. 13, no. 16, p. 2633, 2023.

[114] H. Çetiner and S. Metlek, DenseUNet+: A novel hybrid segmentation approach based on multi-modality images for brain tumor segmentation, Journal of King Saud University - Computer and Information Sciences, vol. 35, no. 8, p. 101663, 2023.

[115] L. Weninger, O. Rippel, S. Koppers, and D. Merhof, Segmentation of Brain Tumors and Patient Survival Prediction: Methods for the BraTS 2018 Challenge, in Crimi, A., Bakas, S., Kuijf, H., Keyvan, F., Reyes, M., van Walsum, T. (eds) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2018. Lecture Notes in Computer Science, vol. 11384, pp. 3–12, Springer, Cham, 2019.

[116] H. Dong, G. Yang, F. Liu, Y. Mo, and Y. Guo, Automatic Brain Tumor Detection and Segmentation Using U-Net Based Fully Convolutional Networks, in Valdés Hernández, M., González-Castro, V. (eds) Medical Image Understanding and Analysis. MIUA 2017. Communications in Computer and Information Science, vol. 723, pp. 506–517, Springer, Cham, 2017.

[117] A. Kermi, I. Mahmoudi, and M. T. Khadir, Deep Convolutional Neural Networks Using U-Net for Automatic Brain Tumor Segmentation in Multimodal MRI Volumes, in Crimi, A., Bakas, S., Kuijf, H., Keyvan, F., Reyes, M., van Walsum, T. (eds) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2018. Lecture Notes in Computer Science, vol. 11384, pp. 37–48, Springer, Cham, 2019.

[118] A. Albiol, A. Albiol, and F. Albiol, Extending 2D Deep Learning Architectures to 3D Image Segmentation Problems, in Crimi, A., Bakas, S., Kuijf, H., Keyvan, F., Reyes, M., van Walsum, T. (eds) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2018. Lecture Notes in Computer Science, vol. 11384, pp. 73–82, Springer, Cham, 2019.

[119] M. U. Rehman, S. Cho, J. Kim, and K. T. Chong, BrainSeg-Net: Brain Tumor MR Image Segmentation via Enhanced Encoder–Decoder Network, Diagnostics, vol. 11, no. 2, p. 169, 2021.

[120] G. Wang, W. Li, S. Ourselin, and T. Vercauteren, Automatic Brain Tumor Segmentation Using Convolutional Neural Networks with Test-Time Augmentation, in Crimi, A., Bakas, S., Kuijf, H., Keyvan, F., Reyes, M., van Walsum, T. (eds) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2018. Lecture Notes in Computer Science, vol. 11384, pp. 61–72, Springer, Cham, 2019.

[121] F. Ullah, S. U. Ansari, M. Hanif, M. A. Ayari, M. E. H. Chowdhury, A. A. Khandakar, and M. S. Khan, Brain MR Image Enhancement for Tumor Segmentation Using 3D U-Net, Sensors, vol. 21, no. 22, p. 7528, 2021.

[122] C. Zhou, S. Chen, C. Ding, and D. Tao, Learning Contextual and Attentive Information for Brain Tumor Segmentation, in Crimi, A., Bakas, S., Kuijf, H., Keyvan, F., Reyes, M., van Walsum,

T. (eds) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2018. Lecture Notes in Computer Science(), vol 11384, vol. 11384, pp. 497–507, Springer, Cham, 2019.

[123] A. Myronenko, 3D MRI Brain Tumor Segmentation Using Autoencoder Regularization, in Crimi, A., Bakas, S., Kuijf, H., Keyvan, F., Reyes, M., van Walsum, T. (eds) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2018. Lecture Notes in Computer Science, vol. 11384, pp. 311–320, Springer, Cham, 2019.

[124] W. Wu, D. Li, J. Du, X. Gao, W. Gu, F. Zhao, X. Feng, and H. Yan, An Intelligent Diagnosis Method of Brain MRI Tumor Segmentation Using Deep Convolutional Neural Network and SVM Algorithm, Computational and Mathematical Methods in Medicine, vol. 2020, pp. 1–10, 2020.

[125] S. Saifullah, A. Pranolo, and R. Dreżewski, Comparative Analysis of Image Enhancement Techniques for Brain Tumor Segmentation: Contrast, Histogram, and Hybrid Approaches, E3S Web Conf., vol. 501, no. International Conference on Computer Science Electronics and Information (ICCSEI 2023), p. 1020, 2024.

[126] F. Isensee, J. Petersen, A. Klein, D. Zimmerer, P. F. Jaeger, S. Kohl, J. Wasserthal, G. Koehler, T. Norajitra, S. Wirkert, and K. H. Maier-Hein, nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation, in Handels, H., Deserno, T., Maier, A., Maier-Hein, K., Palm, C., Tolxdorff, T. (eds) Bildverarbeitung für die Medizin 2019. Informatik aktuell, pp. 22–22, Springer Vieweg, Wiesbaden, 2019.

[127] M. Clerc and J. Kennedy, The particle swarm - explosion, stability, and convergence in a multidimensional complex space, IEEE Transactions on Evolutionary Computation, vol. 6, no. 1, pp. 58–73, 2002.

[128] J. Hu, L. Shen, and G. Sun, Squeeze-and-Excitation Networks, in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7132–7141, IEEE, jun 2018.

[129] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, Attention U-Net: Learning Where to Look for the Pancreas, in Medical Imaging with Deep Learning, apr 2018.

Shoffan Saifullah received a Bachelor's Degree in Informatics Engineering from Universitas Teknologi Yogyakarta, Indonesia, in 2015 and a Master's Degree in Computer Science from Universitas Ahmad Dahlan, Yogyakarta, Indonesia, in 2018. He is currently a lecturer at Universitas Pembangunan Nasional "Veteran" Yogyakarta, Indonesia. His research interests include image processing, computer vision, and artificial intelligence. He is also a Ph.D. student at AGH University of Krakow, Poland, focusing on artificial intelligence (bio-inspired algorithms), image processing, and medical image analysis. He has received grants from AGH University of Krakow and UPN Veteran Yogyakarta, and he is involved in several international collaborative research projects.
https://orcid.org/0000-0001-6799-3834

Rafał Dreżewski received the M.Sc., Ph.D., and D.Sc. (Habilitation) degrees in computer science from the AGH University of Krakow, Poland in 1998, 2005, and 2019, respectively. Since 2019, he has been an Associate Professor with the Institute of Computer Science, AGH University of Krakow, Poland. He is the author of more than 100 papers. His research interests include bio-inspired artificial intelligence algorithms and agent-based modeling and simulation of complex and emergent phenomena.
https://orcid.org/0000-0001-8607-3478

Anton Yudhana received the B.Eng. and M.Eng. degrees in Electrical Engineering from Institut Teknologi Sepuluh Nopember and Universitas Gajah Mada, Indonesia, in 2001 and 2005, respectively, and the Ph.D. degree from Universiti Teknologi Malaysia in 2011. He is an Associate Professor in the Electrical Engineering Department at Universitas Ahmad Dahlan, Indonesia, and the leader of the Center for Electrical and Electronic Research and Development (CEERD). His current research interests include Precision Agriculture, Biomedical Engineering, and Food Engineering based on the Internet of Things.
https://orcid.org/0000-0001-5451-454X

Wahyu Caesarendra received a Bachelor of Engineering (Mechanical) degree from Diponegoro University, Indonesia in 2005. He received New University for Regional Innovation (NURI) and Brain Korea 21 (BK21) scholarships for Master study in 2008 and obtained his Master of Engineering (Mechanical) degree from Pukyong National University, South Korea in 2010. In 2011, Wahyu Caesarendra was awarded of University Postgraduate Award (UPA) and International Postgraduate Tuition Award (IPTA) from the University of Wollongong. He received a Doctor of Philosophy (Mechanical) degree from the University of Wollongong in 2015. He worked as Postdoctoral Research Fellow in Rolls-Royce@NTU Corp Lab,

School of Mechanical and Aerospace Engineering, Nanyang Technological University (NTU), Singapore from February 2017 to September 2018. From October 2018 to October 2024 he served as an Assistant Professor in the Faculty of Integrated Technologies, Universiti Brunei Darussalam. He was a Visiting Assistant Professor at the National Taiwan University of Science and Technology (NTUST) from August 5-11, 2019. He received a funding from Polish National Agency for Academic Exchange (NAWA) and he was a Visiting Academic at Faculty of Mechanical Engineering, Politechnika Opolska from July to December 2023. He was also Visiting Professor in the Department of Mechanical Design Engineering, Vocational School, Diponegoro University in December 2024. From January 2025 to June 2026, he served as a Senior Lecturer in the Department of Mechanical and Mechatronic Engineering at Curtin University Malaysia. He was an Adjunct Professor at the Department of Electrical Automation Engineering, Faculty of Vocation, Institut Teknologi Sepuluh Nopember from January to July 2026. He is currently an Associate Professor at the Faculty of Engineering Technology and Science, Higher Colleges of Technology, United Arab Emirates.

https://orcid.org/0000-0002-9784-4204

Enny Itje Sela received the Doctoral degree in computer science from Universitas Gadjah Mada, Yogyakarta, Indonesia, specializing in medical image processing and voice recognition.She is currently a Professor with the Information Technology Master's Program, Universitas Teknologi Yogyakarta, Yogyakarta, Indonesia. She has been actively involved in research and has received multiple research grants from the Ministry of Education, Culture, Research, and Technology of Indonesia and the National Research and Innovation Agency (BRIN). Her research interests include artificial intelligence, image processing, natural language processing, and intelligent recommender systems. Prof. Sela has been involved in various academic and research collaborations in the field of artificial intelligence and data science. She actively contributes to research and academic publications related to intelligent systems and computational methods.

https://orcid.org/0000-0001-9204-995X