

# A fusion of cross-shaped window attention block and enhanced 3D U-Net for brain tumor segmentation

Ramya Polaki<sup>1</sup>, Prasanna Kumar Rangarajan<sup>1</sup>, Gundala Pallavi<sup>1</sup>, Elakiyya Rajasekhar<sup>2</sup>, Ali Altalbe<sup>3</sup>

<sup>1</sup>Department of Computer Science and Engineering, Amrita Vishwa Vidyapeetham, Chennai, India

<sup>2</sup>Department of Computer Science, Birla Institute of Technology and Science, Dubai, United Arab Emirates

<sup>3</sup>Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

## Article Info

### Article history:

Received May 7, 2024

Revised Jul 30, 2024

Accepted Aug 6, 2024

### Keywords:

3D U-Net

Brain tumors

CrossShaped window attention

Deep learning

Medical imaging

## ABSTRACT

Brain tumor diagnosis and treatment are primarily reliant on medical imaging, necessitating precise segmentation methodologies for practical clinical solutions. Tumor boundaries are difficult to consistently identify, even with breakthroughs in deep learning. To address this challenge, we propose a novel approach that combines an upgraded 3D U-Net architecture for brain tumor segmentation with cross-shaped window attention (CSWA-U-Net). Current segmentation techniques have limitations, particularly in capturing amorphous tumor shapes and fuzzy boundaries. Our strategy aims to overcome these constraints by combining the complementary capabilities of the expanded 3D U-Net, which is efficient at managing volumetric data and maintaining spatial features, with the cross-shaped window attention, which is well-known for capturing long-range relationships and contextual information. We evaluate our method's efficacy using a variety of performance measures, including specificity, sensitivity, and the Dice score. Our results demonstrate increased performance, with Dice scores of 94.7% for the whole tumor, 93.4% for the enhanced tumor region, and 90.5% for the tumor core. Furthermore, our technique has high sensitivity and specificity, highlighting its potential for improving medical imaging analysis.

*This is an open access article under the [CC BY-SA](#) license.*



## Corresponding Author:

Ali Altalbe

Faculty of Computing and Information Technology, King Abdulaziz University

Jeddah 80258, Saudi Arabia

Email: A.altalbe@psau.edu.sa

## 1. INTRODUCTION

Brain tumors result from abnormal cell proliferation, they significantly increase the global mortality rates for both adults and children [1]. The wide array of brain tumors, exceeding 150 types, presents a classification challenge due to their distinct origins within intracranial tissues [2]. Categorizing them into cancerous and noncancerous types is complicated by their individual biology, treatment trajectories, and prognostic factors.

The growing number of brain tumors is probably caused by multiple factors. This tendency can be attributed to various variables such as genetic predisposition, aging populations, potential environmental exposures, diagnostic technological developments, and lifestyle choices. Better imaging methods allow for earlier and more precise identification, but aging populations and possible environmental exposures might make people more susceptible.

Traditional methods for detecting brain tumors have many shortcomings that may affect the accuracy and validity of the diagnosis. Inter-observer variability and impracticality in the event result from the subjectivity and time-consuming nature of manual segmentation and region of interest (ROI) selection. In

addition, differences in detection images used by different observers may cause inconsistent results and affect the accuracy of tumor margin identification. In addition, manual feature extraction may not adequately capture different features of tumor cells, reducing the sensitivity of detection [3]. Conventional methods may encounter problems with visual changes and artifacts and detect small tumors. To overcome these problems, the researchers used methods such as deep learning algorithms that can change and improve the diagnosis of brain diseases. These algorithms are capable of increasing accuracy and efficiency when analyzing large volumes of data and improving patient outcomes.

The ROIs help in tumor characterization, growth assessment, and treatment response evaluation by offering useful information for quantitative analysis. The quality and consistency of the results can be impacted by the subjectivity and inter-observer variability of manual ROI segmentation. Deep learning algorithms are one of the automatic and semi-automatic segmentation techniques [4] that are being developed as alternatives to increase effectiveness and reduce subjectivity in brain tumor ROI segmentation.

Several innovative approaches have been proposed for brain tumor segmentation from magnetic resonance imaging (MRI). MBANet [5], leveraging the BraTS 2018 and 2019 datasets, employs a 3D convolutional neural network with multi-branch attention, achieving superior Dice scores compared to conventional techniques, notably due to the integration of cutting-edge modules like 3D Shuffle Attention. Similarly, DPAFNet [6] introduces a combination of dual-path (DP) and multi-scale attention fusion (MAF) modules alongside 3D feature extraction blocks with residual links and a 3D IDCM module, enhancing context awareness and competitiveness in BraTS2019 Dice scores. Additionally, Ranjbarzadeh *et al.* [7] introduced a novel framework, integrating data from multiple MRI modalities and employing the improved chimp optimization algorithm, showcases advancements in brain tumor segmentation by utilizing support vector machine (SVM) for feature selection and addressing overfitting through data balancing techniques. Furthermore, a patch-based convolutional neural network (CNN) [8] architecture demonstrates robustness and accuracy in segmentation by leveraging convolutional layers for spatially invariant feature learning and integrating various input modalities.

GMetaNet [9], incorporating MetaFormer decoding and a 3D lightweight Ghost CNN, introduces innovative modules for multi-scale feature capture, while AD-Net [10] addresses multimodal feature extraction challenges through efficient channel feature separation learning and regularization techniques using Jensen-Shannon divergence. Finally, a hierarchical multi-view convolution technique, proposed by Guan *et al.* [11], enhances brain tumor segmentation accuracy by obtaining complementary features through decoupling 3D convolution into axial, coronal, and sagittal perspectives, while ensuring parameter consistency via a multi-branch kernel-sharing system with dilated rates.

Montaha *et al.* [12] introduced the use of a 2D U-Net architecture on the BraTS2020 dataset, achieving 99.41% accuracy and 93% dice similarity coefficient (DSC) on T1 MRI sequences. The use of single slices, ablation experiments, and sequence validation improves the robustness and consistency of performance. Furthermore, Edge U-Net model [13], which emphasizes boundary information in MRI images, leads to improved segmentation accuracy through the incorporation of an EGB module and contrast enhancement techniques like CLAHE. Additionally, Aboussaleh *et al.* [14] developed Inception U-Net, a refined U-Net design leveraging inception blocks to enhance performance across multiple BraTS datasets. Subsequently, a multi-input UNet model [15] and scale-wise global contextual axile reverse attention network (SGCARANet) [16], each addressing specific challenges in brain tumor segmentation. These methodologies integrated advanced techniques such as integrated blocks, multiscale dilated features, and attention mechanisms to enhance segmentation accuracy and robustness. MDFU-Net [17], is a unique technique that integrates multiscale dilated features (MDF) and manages heterogeneous data for accurate brain tumor segmentation. This method uses a new Upblock and C-block in the decoder for segmenting and upscaling spatial features, as well as an encoder module using Atr-blocks from deepLabV3+ for multiscale contextual feature extraction.

Continuing their efforts to advance brain tumor segmentation, Jenisha and Shiniha [18] introduced the 3D UNet++ model, a lightweight pseudo-3D architecture combining dense skip connections with 3D convolutions. Metlek and Çetiner [19] presented a novel pre-processing technique focusing on improving tumor visibility in multi-modal images, incorporating a hybrid method with residual blocks to enhance UNet model performance on fine-detail images. Moreover, they demonstrated the effectiveness of focusing solely on the ROI for segmentation tasks.

First of all, CKD-TransBTS [20] presents a clinical knowledge driven model that employs a dual-branch hybrid encoder with modality-correlated cross-attention blocks to reorganize input modalities in accordance with MRI principles. To close the gap between the transformer [21] and CNN features in the decoder [22], it also incorporates a trans and CNN feature calibration block. Second, Datta and Rohilla [23] propose a brain tumor detection and segmentation method that improves feature extraction both locally and globally by utilizing standalone GANs and a vision transformer (ViT) to create artificial images. Last but not

least, AugTransU-Net [24] improves feature variety preservation in transformer-based U-Net models by adding enhanced shortcuts inside these modules and connected attention modules for long-range spatial and channel linkages. Together, these approaches improve brain tumor segmentation by tackling problems with feature extraction, data availability, and processing effectiveness.

Several approaches using Swin transformers have also been proposed with each one providing unique enhancements. BTSwin-Unet presented by Liang *et al.* [25], a 3D U-shaped Swin transformer-based network with self-supervised learning integrated for model pre-training. By utilizing ViTs' multi-head self-attention mechanism, Ghazouani *et al.* [26] suggested a Swin transformer-based network designed for semantic brain tumor segmentation. Zhang *et al.* [27] present IMS2Trans, a scalable and lightweight Swin transformer network that addresses incomplete modalities and multi-modal MRI data processing effectively. To extract features for all observable modalities, this technique uses a single encoder with common weights. It also uses a feature distillation strategy for consistency regularization, which is then aggregated in the decoder. The last contribution from ZongRen *et al.* [28] is DenseTrans, which incorporates Swin transformer into the UNet++ architecture for brain tumor segmentation. It emphasizes the use of control mechanisms and deep separable convolution to strike a compromise between computing complexity and accuracy.

With a reported frequency of 5-10 occurrences per 100,000 persons, brain tumors are a serious health risk in India, accounting for over 28,000 cases and over 24,000 fatalities annually [29]. Brain tumors can be generically categorized as benign, which usually grows slowly with defined borders, or malignant, which is characterized by uncontrollably growing and spreading. The primary objectives of this research paper that focus on contributing to the brain tumor segmentation paradigm are:

- Development of a new medical imaging technique CSWA-UNet by combining an enhanced 3D U-Net with cross-shaped window attention to accurately segment brain tumors.
- Validate the integrated model's performance in precisely identifying the borders of brain tumors in medical images.

The paper is organized systematically. Section 2 describes the study's methodology, focusing on the dataset, core framework, and approaches. Sections 3 and 4 contain the experimental results and a detailed analysis of the model's performance. Section 5 finishes the work by summarizing its contributions and proposing future research possibilities.

## 2. METHOD

Research on brain tumor segmentation uses a dataset called BraTS 2020 (brain tumor segmentation), which is a collection of multimodal MRI images in NIFTI file format (.nii.gz). It consists of T1-weighted, T1-weighted contrast-enhanced, T2-weighted, and fluid attenuated inversion recovery (FLAIR) volumes that were acquired from different clinical regimens and scanners at different institutions as shown in Figure 1. One to four raters followed the same technique and competent neuro-radiologists authorized the manual segmentations of GD-enhancing tumor (ET), peritumoral edema (ED), and necrotic and non-enhancing tumor core (NCR/NET) annotations that are included in the dataset as shown in Figure 2.

These annotations are made available after pre-processing procedures such as skull-stripping, co-registration, and interpolation to a uniform resolution (1 mm<sup>3</sup>). There are 369 folders in the BraTS 2020 Training Data directory, each of which represents a different dataset case. Essential MRI modalities are in these folders and supplied as NIFTI files (.nii). These modalities provide thorough imaging data that is essential for precise segmentation and analysis of brain tumors. In addition, every case has a segmentation mask file called "seg.nii" that has annotations that show the locations of the tumors.

Two CSV files, "Namemapping.csv" and "survivalinfo.csv," which provide further information and data annotations, are also included in the directory. When combined, these resources provide researchers with important information for improving algorithms and techniques related to the diagnosis and planning of brain tumors. The preprocessing procedures include importing the MRI images and related segmentation masks, cropping them to a consistent size, normalizing the image data, and preprocessing the masks to designate various tumor locations. During training, augmenting techniques such as rotation and flipping are used. Finally, the data is returned as tensors, ready to be trained.

Data mapping and patient survival statistics for brain tumors are loaded and combined from CSV files to start the preprocessing process. Paths to specific patient data files are then created based on the patients' IDs and whether or not they are part of the testing or training stages. Next, using stratified Kfold cross-validation with K value 7, the dataset is divided into training and validation subsets to guarantee proportionate representation across age groups.

Segmentation masks are used to analyze brain tumors. As shown in Figure 3 divides the original mask into three binary masks that represent different tumor regions: the whole tumor (WT), the tumor core (TC), and the enhancing tumor (ET). The WT mask includes all important tumor regions. Still, the TC mask concentrates on the necrotic/non-enhancing and enhancing core regions, and the ET mask isolates the

enhancing core region. These masks are then integrated into a multi-channel mask, with each channel representing a distinct tumor location. This pre-processing simplifies the segmentation effort by categorizing tumor regions and preparing masks for use in neural network models, allowing for more accurate segmentation of various tumor sub-regions.

The process starts with data preprocessing as demonstrated in the Figure 4, in which MRI images and segmentation masks are loaded, normalized, and augmented to increase the dataset's diversity. Stratified K-Fold cross-validation ensures a balanced distribution across folds, allowing for more robust model evaluation. The modified U-Net with cross-shaped window attention serves as the pipeline's heart. In the network's forward pass, the input is a feature extracted via the context route, which extracts hierarchical features at several scales. Then, a cross-shaped window attention method is used to improve the informative content of the features retrieved from the context pathway. These increased features are subsequently combined with features from the localization route, which is in charge of reconstructing spatial resolution.

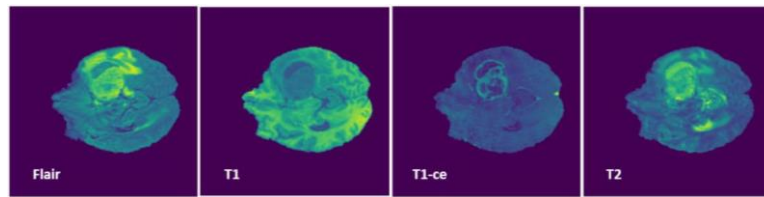


Figure 1. Multi-modal MRI: FLAIR, T1, T1-CE, and T2 volumes for detailed tissue analysis

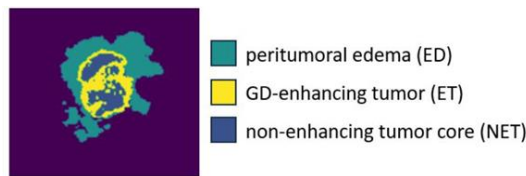


Figure 2. Segmentation mask with annotated labels

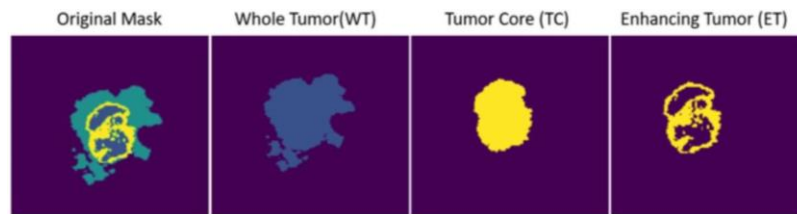


Figure 3. Visualization of whole tumor (WT), tumor core (TC), and enhancing tumor (ET)

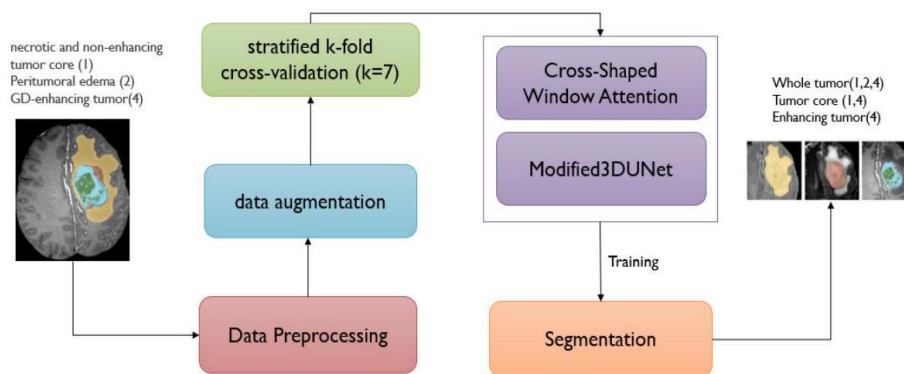


Figure 4. Brain tumor segmentation algorithm

This concatenated feature representation is subsequently processed to yield the final segmentation output. Throughout the network, residual connections are used to handle the vanishing gradient problem, in which a layer's input is added to its output via element-wise summing, allowing for smoother gradient flow during training and preserving fine-grained spatial features.

### 2.1. Context pathway

The context pathway in the model architecture is divided into layers as seen in Figure 5, each designed to extract hierarchical aspects from the input data while gradually widening the receptive area. Initially, a 3D convolutional operation is used to keep the input size constant as shown in Algorithm 1, followed by another convolutional layer with a stride of 2 to collect more context. Instance normalization and LeakyReLU activation stabilize training while introducing nonlinearity. Subsequent levels double the number of filters to improve feature representation, with downsampling only used at the initial level to retain spatial resolution. Furthermore, residual connections improve information flow and alleviate the vanishing gradient problem.

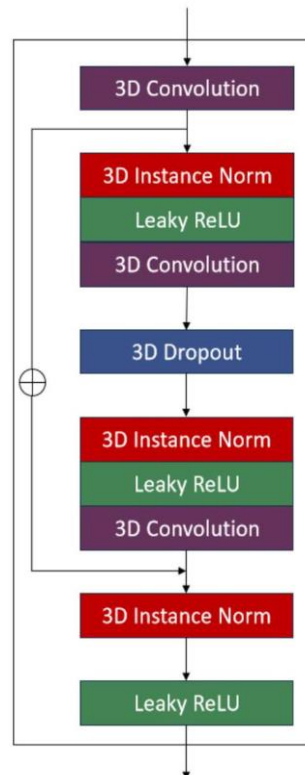


Figure 5. Context pathway block diagram

#### Algorithm 1. Context pathway algorithm

```

 $x_{conv1} \leftarrow \text{Conv}(x, W_1) + b_1$ 
 $x_{norm1} \leftarrow \text{ReLU}(\text{Norm}(x_{conv1}))$ 
 $x_{conv2} \leftarrow \text{Conv}(x_{norm1}, W_2) + b_2$ 
 $x_{norm2} \leftarrow \text{ReLU}(\text{Norm}(x_{conv2}))$ 
 $x_{conv3} \leftarrow \text{Conv}(x_{norm2}, W_3) + b_3$ 
 $x_{residual} \leftarrow x_{conv3} + x_{norm2}$ 
 $x_{norm3} \leftarrow \text{ReLU}(\text{Norm}(x_{residual}))$ 
return  $x_{norm3}$ 

```

Furthermore, 3D instances normalization enables consistent training dynamics following each convolutional operation. The initial layer's output feeds subsequent layers and a cross-shaped window attention mechanism, which improves features by capturing spatial dependencies. By incorporating insights from the following layers and the attention mechanism [30], [31], the model successfully captures contextual information while focusing on relevant features, increasing performance in tasks such as segmentation and classification.

## 2.2. Cross-shaped window attention layer

Cross-shaped windows self-attention (CSWA) is a novel attention mechanism developed for processing 3D data in convolutional neural networks (CNNs) [32] as shown in Figure 6. It presents a novel method for capturing spatial interdependence within the input volume by splitting it into horizontal, vertical, and longitudinal stripes and paying attention to each stripe independently. The attention layer subdivides the input into  $G$  heads. The initial  $G/3$  heads concentrate on horizontal attention, the following  $G/3$  on vertical attention, and the remaining heads on longitudinal attention. Each attention mechanism operates on a subset of the input tensor defined by the strip width parameter ( $sw$ ). Horizontal attention focuses on regions of  $sw \times W \times D$ , vertical attention on  $sw \times H \times D$  and longitudinal attention on  $sw \times H \times W$ . This structured technique allows the model to efficiently capture information from several dimensions of the input tensor.

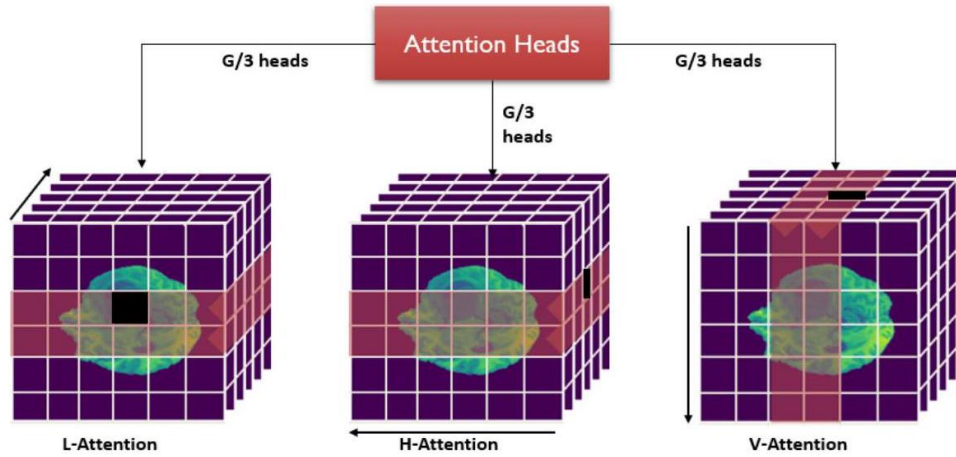


Figure 6. Cross-shaped window attention layer

Suppose the projected queries (Q), keys (K), and values (V) of the  $k$ -th head all have dimension  $dk$ , then the output of the horizontal, vertical, and longitudinal stripes self-attention for  $k$ -th head is defined as (1)-(5):

$$T = [P_{h1}, \dots, P_{hM}, P_{v1}, \dots, P_{vM}, P_{l1}, \dots, P_{lM}] \quad (1)$$

$$A_{ki} = \text{Attention}(P_i W_k^Q, P_i W_k^K, P_i W_k^V) \quad (2)$$

$$H\text{Attention}_k(X) = [P_k^1, P_k^2, \dots, P_k^M], \text{ for } k = 1, 2, \dots, \frac{G}{3} \quad (3)$$

$$V\text{Attention}_k(X) = [P_k^1, P_k^2, \dots, P_k^M], \text{ for } k = \frac{G}{3} + 1, \dots, 2\frac{G}{3} \quad (4)$$

$$L\text{Attention}_k(X) = [P_k^1, P_k^2, \dots, P_k^M], \text{ for } k = 2\frac{G}{3} + 1, \dots, G \quad (5)$$

In CSWA,  $A_{ki}$  denotes the attention scores for the  $i$ -th stripe in the  $k$ -th head, computed by the attention () mechanism using projected queries, keys, and values represented as  $P_i W_k^Q$ ,  $P_i W_k^K$ , and  $P_i W_k^V$  correspondingly. The outputs of the horizontal, vertical, and longitudinal attention mechanisms for the  $k$ -th head is denoted by  $H\text{Attention}_k(X)$ ,  $V\text{Attention}_k(X)$ , and  $L\text{Attention}_k(X)$ , respectively. These outputs are concatenated using the  $\text{Concat}()$  method to produce the final CSWA output, indicated as  $CSWA(X)$ . This approach effectively captures spatial dependency in 3D data by focusing on distinct parts of the input tensor in several dimensions.

## 2.3. Localization pathway

A U-Net architecture's localization pathway as shown in Figure 7 receives feature maps from the corresponding level in the encoding pathway, which are concatenated with feature maps from the previous level in the localization pathway. Furthermore, at each stage of the localization route, the feature maps are

supplemented with refined features obtained via the cross-shaped windows attention method. This integration enables the model to incorporate spatial dependencies detected by the attention mechanism, hence improving the localization pathway's ability to refine and recover spatial features while maintaining semantic information.

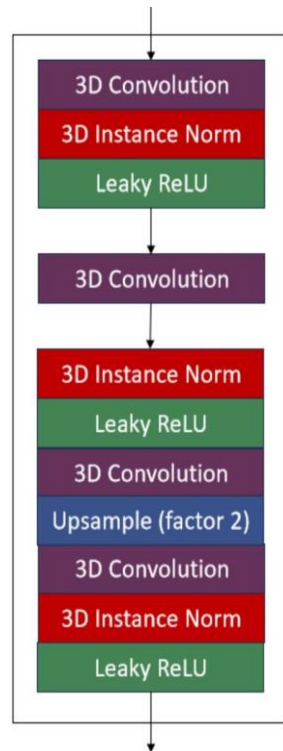


Figure 7. Localization pathway block diagram

The concatenated feature maps are then subjected to convolutional processes before being normalized, which standardizes the features and helps to stabilize the learning process. In addition, the Leaky ReLU activation function is used to introduce non-linearity, allowing the model to detect complicated patterns in the data. Following that, the feature maps are upsampled to boost spatial resolution while preserving fine-grained features. Another convolutional layer is then used to refine the upsampled features. Finally, instance normalization and Leaky ReLU activation are used again to ensure consistency and non-linearity in the revised feature representations as shown in Algorithm 2. This repeating procedure at various stages of the localization route helps to generate accurate segmentation predictions by effectively recovering spatial details and semantic information from encoded characteristics.

#### Algorithm 2. Localization pathway algorithm

```

 $x_{conv} \leftarrow \text{Conv}(x, W_{conv}) + b_{conv}$ 
 $x_{norm} \leftarrow \text{Norm}(x_{conv})$ 
 $x_{relu} \leftarrow \text{ReLU}(x_{norm})$ 
 $x_{upsample} \leftarrow \text{Upsample}(x_{relu}, W_{upsample}) + b_{upsample}$ 
 $x_{conv2} \leftarrow \text{Conv}(x_{upsample}, W_{conv}) + b_{conv}$ 
 $x_{norm2} \leftarrow \text{Norm}(x_{conv2})$ 
 $x_{relu2} \leftarrow \text{ReLU}(x_{norm2})$ 
return  $x_{relu2}$ 

```

During training, the model shown in the Figure 8 processes batches of medical pictures and masks across numerous epochs. It starts with a forward pass that generates predictions from the input images, and then calculates the loss by comparing the predictions to the ground truth masks. Gradients are then calculated via backpropagation, allowing the model parameters to be changed with the Adam optimizer. The learning rate may alter dynamically in response to validation loss trends. The model periodically stores its state, or checkpoint, to keep the best-performing version depending on validation results. After training, the model can create predictions for fresh pictures, and their accuracy is measured using evaluation metrics such as Dice score, sensitivity, and specificity, which provide information about segmentation accuracy.

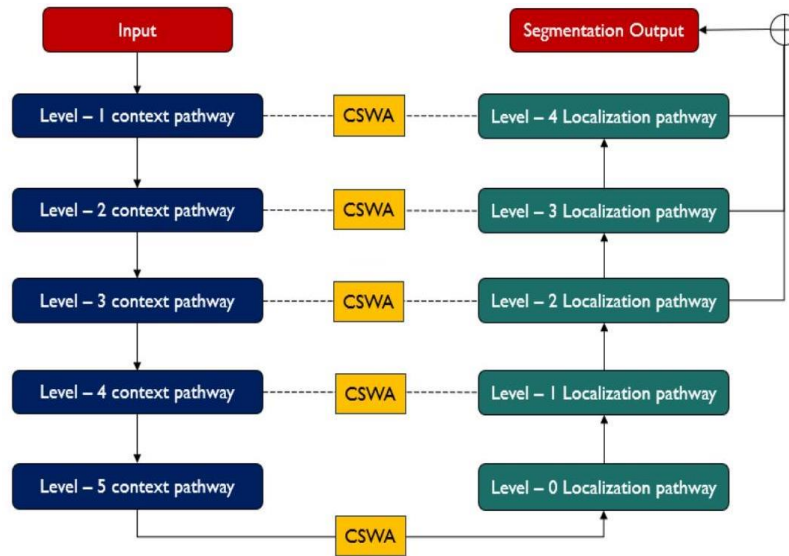


Figure 8. Cross-shaped window attention (CSWA) block and enhanced 3D U-Net for brain tumor segmentation

### 3. RESULTS AND DISCUSSION

The performance of the trained model is evaluated using a variety of measures and compared to previous methodologies or baseline data. The section begins by showing the quantitative evaluation metrics from the validation dataset, such as the Dice score, sensitivity, and specificity for each class (*e.g.*, total tumor, tumor core, and enhanced tumor). These measurements provide insight into the model's segmentation accuracy across various tumor regions.

Additionally, qualitative assessments can be supplied by overlaying model predictions on the source photos or ground truth masks. This gives a more intuitive sense of how well the model reflects tumor boundaries and regions of interest. We employed the Tesla T4, equipped with 16 GB of GDDR6 memory and boasting 2,560 CUDA cores, for the computational tasks. Table 1 here shows the hyperparameters required for the training purpose.

Table 1. Hyperparameters for training

Parameter	Value
Number of workers for data loader	4
Batch size during training	32
Spatial size of training images	64
Number of channels in the training images	3
Number of training epochs	50
The learning rate for optimizers	0.0005

In image segmentation tasks, Dice scores are used as a statistic to quantify the similarity between anticipated and ground truth segmentations. The formula for the Dice score is stated as:

$$\text{Dice} = \frac{2 \times |A \cap B|}{|A| + |B|}$$

where  $A$  represents the set of pixels classified as positive in the ground truth,  $B$  represents the set of pixels classified as positive in the prediction, and  $|A \cap B|$  denotes the cardinality of a set. A Dice score of 1 indicates perfect overlap between the prediction and ground truth, while a score of 0 indicates no overlap.

The binary cross-entropy+dice loss (BCEDice) loss [33] is a composite loss function used for training image segmentation models. To optimize model parameters, it uses both binary cross-entropy (BCE) loss and Dice Loss. BCE Loss measures the dissimilarity between predicted probabilities and ground truth labels for individual pixels. Meanwhile, Dice Loss calculates dissimilarity using the overlap between expected and ground truth segmentations. By combining BCE Loss and Dice Loss, BCEDice Loss promotes accurate pixel classification and a significant overlap with ground truth. The BCEDice Loss formula is:



$$BCEDice Loss = BCE Loss + Dice Loss$$

After analyzing the presented coefficients as shown in Figure 9, we find that the WT class performs the best across all parameters, with a Dice score of 94.7%, sensitivity of 93.7%, and specificity of 95.2%. The TC class likewise performs well, but slightly below WT, with a Dice score of 94.4%, sensitivity of 94.4%, and specificity of 94.1%. On the other hand, the ET class does relatively poorly, with a Dice score of 90.5%, sensitivity of 89.6%, and specificity of 89.9%.

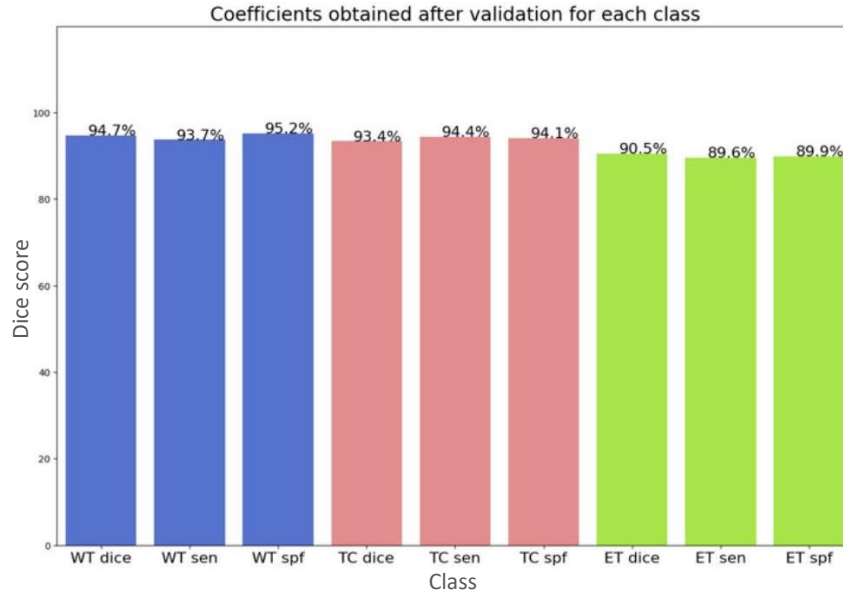


Figure 9. Dice score, sensitivity and specificity for WT, TC, and ET classes

$$BCE Loss = -\frac{1}{N} \sum_{i=1}^N (y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i))$$

Where  $N$  represents the total number of pixels in the image,  $y_i$  denotes the ground truth label for each pixel  $i$ , which can take values of either 0 or 1, indicating the absence or presence of the target class, respectively. Similarly,  $p_i$  signifies the predicted probability assigned to pixel  $i$  by the model, reflecting the likelihood of it belonging to the positive class.

$$Dice Loss = 1 - Dice Score$$

The bar plot Figure 10 shows the Dice scores obtained by three different models - 3DUNet [31], 3D Attention U-Net [31], and our model - in three unique classes: WT, TC, and ET on BraTS 2020 dataset. The height of each bar relates to the Dice score achieved by the model for a specific class. Notably, our model consistently beats the other models in all classes, with the greatest Dice scores for WT, TC, and ET.

This visualization provides vital insights into the models' segmentation performance and demonstrates Our Model's superior performance when compared to existing methodologies. Over 50 epochs in Figure 11 shows the model's performance over 50 epochs, with the loss curve in Figure 11(a) indicating better performance as the values decrease, and the Dice score curve in Figure 11(b) reflecting improved segmentation accuracy with higher values. Here the loss curve demonstrates how well the neural network model matches the training data, with decreasing values indicating better performance. In contrast, the Dice curve measures the model's segmentation accuracy, with higher values indicating greater overlap between anticipated and ground truth segmentations. Ideally, both curves should show decreasing loss and increasing Dice coefficient, suggesting successful training.

As shown in the Figure 12 ground truth mask images exhibit hand-annotated segmentation maps, which serve as the gold standard for evaluating segmentation model performance. In contrast, the model generates predicted mask visuals, which represent its segmentation output based on learning parameters and input data. Comparing the images enabled us to evaluate the model's accuracy in segmenting objects or regions of interest.

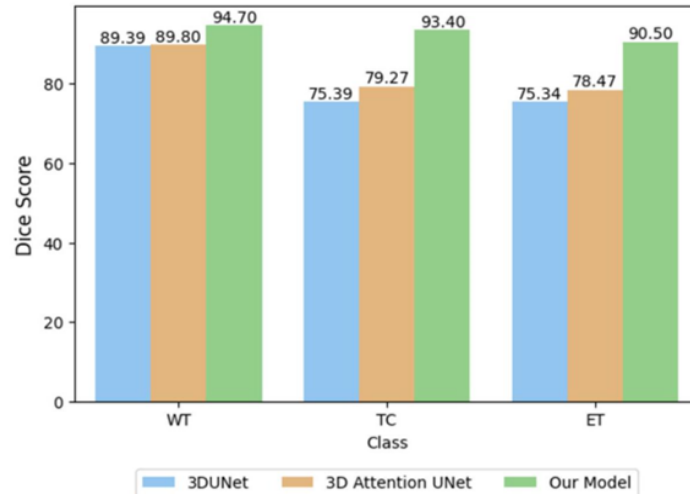


Figure 10. Comparison of 3DUNet, 3D Attention U-Net and our model's Dice score for WT, TC, and ET classes

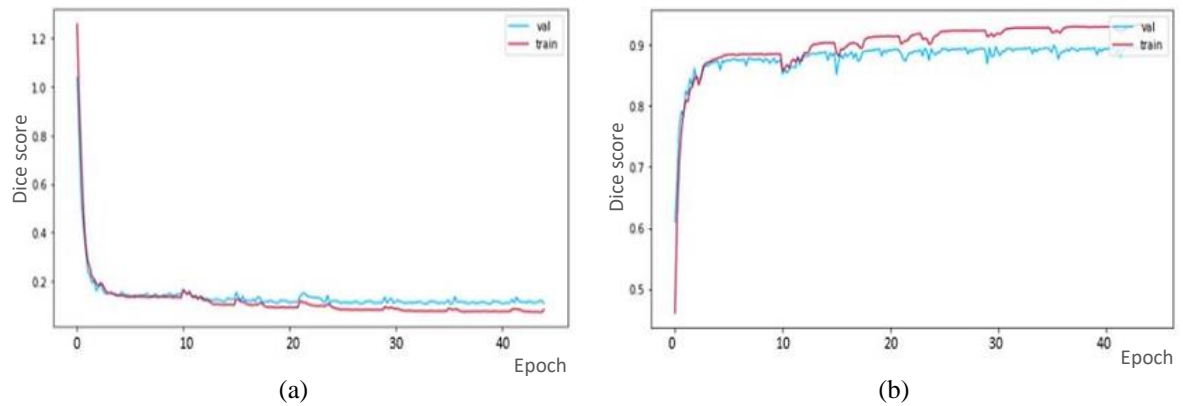


Figure 11. Training and validation performance metrics over 50 epochs: 11(a) loss curve and 11(b) Dice score curve for both training and validation sets

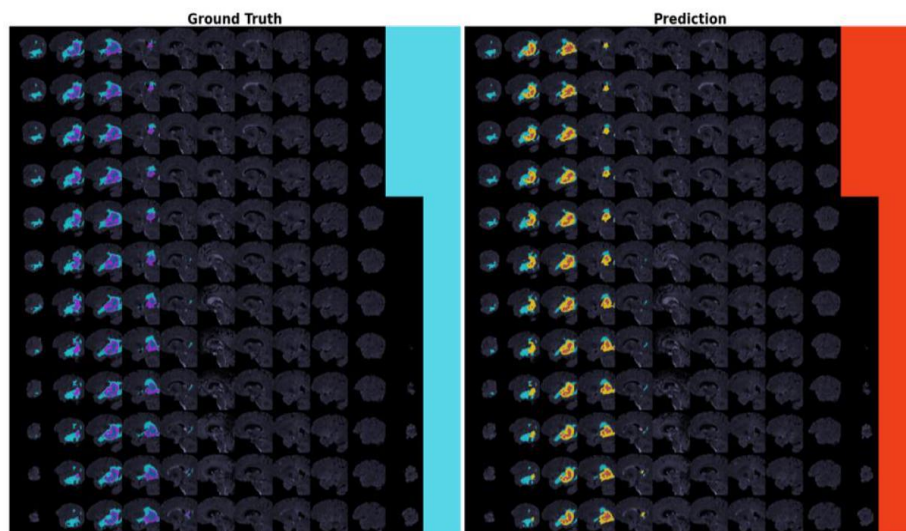


Figure 12. Ground truth and the prediction masks

#### 4. CONCLUSION

In conclusion, segmenting brain tumor subregions is an important work in medical image processing, since it allows for proper diagnosis and therapy planning. This paper presents a novel strategy to doing this task that employs a modified 3D U-Net architecture with a context and localization pathway, as well as CrossShaped window attention. The model outperforms established benchmarks in segmenting tumor subregions, with validation Dice scores of 94.7% for WT, 93.4% for TC, and 90.5% for ET. These findings underscore the model's capacity to precisely identify tumor boundaries, which is critical for clinical decision-making.

For the research field, these findings signify a significant advancement in medical image processing. The model's high performance sets a new benchmark for brain tumor segmentation, encouraging further exploration and refinement of similar approaches. The integration of a context and localization pathway, along with CrossShaped Window Attention, offers a novel direction for future research, potentially applicable to other areas of medical imaging and beyond.

The proposed model's improved feature extraction capabilities provide exceptional precision in identifying tumor subregions, even when data availability is low. In clinical practice, this precision enables healthcare providers to properly diagnose and arrange treatment plans for patients with brain tumors. In the future, the model's performance and computing efficiency could be improved by including 3D imaging techniques and further exploring quantum-inspired approaches, such as employing quantum-based feature extraction methods or reinforcement learning algorithms. Furthermore, examining object segmentation approaches may provide a more detailed understanding of tumor form. By addressing these areas for development, the potential for more precise and accurate brain tumor segmentation models becomes clear.

#### ACKNOWLEDGEMENTS

The authors extend their appreciation to Prince Sattam bin Abdulaziz University for funding this research work through the project number (PSAU/2024/01/823183).




#### REFERENCES

- [1] U. F. M. Muhammad Asif, Muhammad Saleem, "Identification and prediction of brain tumor using VGG-16 empowered with explainable artificial intelligence," *International Journal of Computational and Innovative Sciences*, vol. 2, no. 2, pp. 24–33, 2023.
- [2] R. Ranjbarzadeh, A. Caputo, E. B. Tirkolaei, S. J. Ghouschi, and M. Bendeche, "Brain tumor segmentation of MRI images: a comprehensive review on the application of artificial intelligence tools," *Computers in Biology and Medicine*, vol. 152, Jan. 2023, doi: 10.1016/j.combiomed.2022.106405.
- [3] B. Babu Vimala, S. Srinivasan, S. K. Mathivanan, Mahalakshmi, P. Jayagopal, and G. T. Dalu, "Detection and classification of brain tumor using hybrid deep learning models," *Scientific Reports*, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-50505-6.
- [4] J. Tian, D. Dong, Z. Liu, and J. Wei, "Introduction," in *Radiomics and Its Clinical Application*, Elsevier, 2021, pp. 1–18, doi: 10.1016/B978-0-12-818101-0.00004-5.
- [5] Y. Cao, W. Zhou, M. Zang, D. An, Y. Feng, and B. Yu, "MBANet: a 3D convolutional neural network with multi-branch attention for brain tumor segmentation from MRI images," *Biomedical Signal Processing and Control*, vol. 80, 2023, doi: 10.1016/j.bspc.2022.104296.
- [6] Y. Chang, Z. Zheng, Y. Sun, M. Zhao, Y. Lu, and Y. Zhang, "DPAFNet: a residual dual-path attention-fusion convolutional neural network for multimodal brain tumor segmentation," *Biomedical Signal Processing and Control*, vol. 79, 2023, doi: 10.1016/j.bspc.2022.104037.
- [7] R. Ranjbarzadeh, P. Zarkhsh, A. Caputo, E. B. Tirkolaei, and M. Bendeche, "Brain tumor segmentation based on optimized convolutional neural network and improved chimp optimization algorithm," *Computers in Biology and Medicine*, vol. 168, 2024, doi: 10.1016/j.combiomed.2023.107723.
- [8] F. Ullah, A. Salam, M. Abrar, and F. Amin, "Brain tumor segmentation using a patch-based convolutional neural network: a big data analysis approach," *Mathematics*, vol. 11, no. 7, 2023, doi: 10.3390/math11071635.
- [9] Y. Lu *et al.*, "GMetaNet: multi-scale ghost convolutional neural network with auxiliary MetaFormer decoding path for brain tumor segmentation," *Biomedical Signal Processing and Control*, vol. 83, 2023, doi: 10.1016/j.bspc.2023.104694.
- [10] Y. Peng and J. Sun, "The multimodal MRI brain tumor segmentation based on AD-Net," *Biomedical Signal Processing and Control*, vol. 80, 2023, doi: 10.1016/j.bspc.2022.104336.
- [11] X. Guan, Y. Zhao, C. O. Nyatega, and Q. Li, "Brain tumor segmentation network with multi-view ensemble discrimination and Kernel-Sharing dilated convolution," *Brain Sciences*, vol. 13, no. 4, 2023, doi: 10.3390/brainsci13040650.
- [12] S. Montaha, S. Azam, A. K. M. R. H. Rafid, M. Z. Hasan, and A. Karim, "Brain tumor segmentation from 3D MRI scans using U-Net," *SN Computer Science*, vol. 4, no. 4, 2023, doi: 10.1007/s42979-023-01854-6.
- [13] A. M. Gab Allah, A. M. Sarhan, and N. M. Elshennawy, "Edge U-Net: brain tumor segmentation using MRI based on deep U-Net model with boundary information," *Expert Systems with Applications*, vol. 213, 2023, doi: 10.1016/j.eswa.2022.118833.
- [14] I. Aboussaleh, J. Riffi, A. M. Mahraz, and H. Tairi, "Inception-UDet: an improved U-Net architecture for brain tumor segmentation," *Annals of Data Science*, vol. 11, no. 3, pp. 831–853, 2024, doi: 10.1007/s40745-023-00480-6.
- [15] L. Fang and X. Wang, "Multi-input UNet model based on the integrated block and the aggregation connection for MRI brain tumor segmentation," *Biomedical Signal Processing and Control*, vol. 79, 2023, doi: 10.1016/j.bspc.2022.104027.
- [16] M. Karri, C. S. R. Annvarapu, and U. R. Acharya, "SGC-ARANet: scale-wise global contextual axile reverse attention network for automatic brain tumor segmentation," *Applied Intelligence*, vol. 53, no. 12, pp. 15407–15423, 2023, doi: 10.1007/s10489-022-04209-5.




- [17] H. Sultan *et al.*, “MDFU-Net: multiscale dilated features up-sampling network for accurate segmentation of tumor from heterogeneous brain data,” *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 5, 2023, doi: 10.1016/j.jksuci.2023.101560.
- [18] J. Jenisha and A. M. J. Shiniha, “Efficient brain tumor segmentation using lightweight pseudo-3D UNet++ model,” *EPRA International Journal of Research and Development (IJRD)*, vol. 8, no. 7, pp. 215–220, 2023.
- [19] S. Metlek and H. Çetiner, “ResUNet+: a new convolutional and attention block-based approach for brain tumor segmentation,” *IEEE Access*, vol. 11, pp. 69884–69902, 2023, doi: 10.1109/ACCESS.2023.3294179.
- [20] J. Lin *et al.*, “CKD-TransBTS: clinical knowledge-driven hybrid transformer with modality0correlated cross-attention for brain tumor segmentation,” *IEEE Transactions on Medical Imaging*, vol. 42, no. 8, pp. 2451–2461, Aug. 2023, doi: 10.1109/TMI.2023.3250474.
- [21] B. M. Gurusamy, P. K. Rengarajan, and P. Srinivasan, “A hybrid approach for text summarization using semantic latent Dirichlet allocation and sentence concept mapping with transformer,” *International Journal of Electrical and Computer Engineering*, vol. 13, no. 6, pp. 6663–6672, 2023, doi: 10.11591/ijece.v13i6.pp6663-6672.
- [22] J. A. Prakash *et al.*, “Transfer learning approach for pediatric pneumonia diagnosis using channel attention deep CNN architectures,” *Engineering Applications of Artificial Intelligence*, vol. 123, 2023, doi: 10.1016/j.engappai.2023.106416.
- [23] P. Datta and R. Rohilla, “Brain tumor image pixel segmentation and detection using an aggregation of GAN models with vision transformer,” *International Journal of Imaging Systems and Technology*, vol. 34, no. 1, 2024, doi: 10.1002/ima.22979.
- [24] M. Zhang *et al.*, “Augmented transformer network for MRI brain tumor segmentation,” *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 1, 2024, doi: 10.1016/j.jksuci.2024.101917.
- [25] J. Liang, C. Yang, J. Zhong, and X. Ye, “BTSwin-Unet: 3D U-shaped symmetrical Swin transformer-based network for brain tumor segmentation with self-supervised pre-training,” *Neural Processing Letters*, vol. 55, no. 4, pp. 3695–3713, Aug. 2023, doi: 10.1007/s11063-022-10919-1.
- [26] 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, 2024, doi: 10.1007/s11548-023-03024-8.
- [27] D. Zhang, C. Wang, T. Chen, W. Chen, and Y. Shen, “Scalable Swin transformer network for brain tumor segmentation from incomplete MRI modalities,” *Artificial Intelligence in Medicine*, vol. 149, 2024, doi: 10.1016/j.artmed.2024.102788.
- [28] L. ZongRen, W. Silamu, W. Yuzhen, and W. Zhe, “DenseTrans: multimodal brain tumor segmentation using Swin transformer,” *IEEE Access*, vol. 11, pp. 42895–42908, 2023, doi: 10.1109/ACCESS.2023.3272055.
- [29] Healthworld, “World brain tumour day 2022 ‘together we are stronger,’” *health.economictimes.indiatimes.com*, 2022. <https://health.economictimes.indiatimes.com/> (accessed: Jun. 08, 2022).
- [30] G. Bharathi Mohan, R. Prasanna Kumar, R. Elakkiya, and B. G. Prasanna Kumar, “Medical recommendations: leveraging CRNN with self-attention mechanism for enhanced systems,” in *2023 International Conference on Evolutionary Algorithms and Soft Computing Techniques, EASCT 2023*, 2023, pp. 1–6, doi: 10.1109/EASCT59475.2023.10393094.
- [31] G. Manimaran and J. Swaminathan, “Focal-WNet: an architecture unifying convolution and attention for depth estimation,” in *2022 IEEE 7th International conference for Convergence in Technology, I2CT 2022*, 2022, pp. 1–7, doi: 10.1109/I2CT54291.2022.9824488.
- [32] Y. Li *et al.*, “Cross-shaped windows transformer with self-supervised pretraining for clinically significant prostate cancer detection in bi-parametric MRI,” *arXiv preprint arXiv:2305.00385*, 2023.
- [33] M. Yeung, E. Sala, C.-B. Schönlieb, and L. Rundo, “Unified focal loss: generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation,” *Computerized Medical Imaging and Graphics*, vol. 95, p. 102026, Jan. 2022, doi: 10.1016/j.compmedimag.2021.102026.

## BIOGRAPHIES OF AUTHORS






**Ramya Polaki**    is a dynamic computer science graduate specializing in AI from Amrita Vishwa Vidyapeetham, Chennai. With a focus on data-driven discovery and AI for Healthcare, Ramya has authored two papers indexed in Scopus. Her research encompasses diverse topics, including the analysis of COVID-19 using cough recordings to determine the efficacy of different models. Additionally, she has investigated brain tumor detection using ResNet and particle swarm optimization techniques. Passionate about innovation, Ramya leverages advanced technologies to push boundaries in research and development. She can be reached at email: ramyapolaki6046@gmail.com.






**Prasanna Kumar Rangarajan**    currently serves as the chairperson and associate professor, at the School of Computing, Chennai. With 22 years of experience in the field of teaching, he has established himself as an expert in several areas of computer science. His interests span across data analytics, machine learning, theory of computation, compiler design, and python programming. Currently, Dr. Kumar is supervising 8 Ph.D. research scholars and has successfully guided 8 M.E projects. Additionally, he serves as a Doctoral Committee Member for 10 Ph.D. research scholars. He has been invited as a guest lecturer and resource person for Faculty Development Programs and has also contributed as a program committee member and session chair in various national and international conferences. She can be reached at email: r\_prasannakumar@ch.amrita.edu.






**Gundala Pallavi**    is a dedicated researcher currently pursuing a Ph.D. in computer science and engineering with a specialization in artificial intelligence (AI) at Amrita Vishwa Vidyapeetham, Chennai, India. Her interdisciplinary background spans multiple fields including machine learning (ML), deep learning, quantum natural language processing (NLP), and Bioinformatics. Her work aims to develop novel algorithms and methodologies that leverage the power of quantum computing to revolutionize the field of bioinformatics. She can be reached at [g\\_pallavi@ch.students.amrita.edu](mailto:g_pallavi@ch.students.amrita.edu).



**Elakiyya Rajasekhar**    received the Ph.D. degree from Anna University, Chennai, in 2018. She was an assistant professor with the Department of Computer Science and Engineering, School of Computing, SASTRA University, Thanjavur. She is currently an assistant professor with the Department of Computer Science, Birla Institute of Technology and Science, Pilani, Dubai Campus, Dubai International Academic City Dubai, United Arab Emirates. She has three patents and has published more than 35 research papers in leading journals, conference proceedings, and books, including IEEE, Elsevier, and Springer. Her research interests include deep learning and computer vision. She is a lifetime member of the International Association of Engineers. She is also an editor of the Information Engineering and Applied Computing Journal. She has organized various events, including a Workshop on “Cyber security” with the Agni College of Technology, Chennai, from 2014 to 2017, sponsored by IIT Bombay, and a Software Development Workshop, “EKTIA” with the Jerusalem College of Engineering, Chennai, in 2012. She can be reached at email: [elakkiya@dubai.bits-pilani.ac.in](mailto:elakkiya@dubai.bits-pilani.ac.in).



**Ali Altalbe**    received the M.Sc. degree in information technology from Flinders University, Australia, and the Ph.D. degree in information technology from the University of Queensland, Australia. He is currently an associate professor with the Department of IT, King Abdulaziz University, Jeddah, Saudi Arabia. He can be reached at [A.altalbe@psau.edu.sa](mailto:A.altalbe@psau.edu.sa).