Applied Mathematics & Information Sciences
*An International Journal*

http://dx.doi.org/10.18576/amis/200303

# Enhanced Brain Tumor Segmentation Using Multi-Stream Hybrid Deep Learning with Cross-Attention Integration

*Sulieman Shelash Mohammad*[1,2,*], *Hamza Abu Owida*[3], *Suhaila Abuowaida*[4], *Asokan Vasudevan*[2], *Nawaf Alshdaifat*[5], *and Mohammad Faleh Ahmmad Hunitie*[6]

[1]Electronic Marketing and Social Media, Economic and Administrative Sciences, Zarqa University, Zarqa 13110, Jordan
[2]Faculty of Business and Communications, INTI International University, Negeri Sembilan 71800, Malaysia
[3]Department of Medical Engineering, Faculty of Engineering, Al-Ahliyya Amman University, Ammman, 19328, Jordan
[4]Department of Computer Science, Faculty of Prince Al-Hussein Bin Abdallah II for IT, Al al-Bayt University, Mafraq, Jordan
[5]Department of Information Technology, Faculty of Prince Al-Hussein bin Abdullah II of Information Technology, The Hashemite University, Zarqa, Jordan
[6]Department of Public Administration, School of Business, University of Jordan, Jordan

Received: 7 Jul. 2025, Revised: 28 Dec. 2025, Accepted: 15 Jan. 2026

**Abstract:** This paper proposes a novel hybrid deep learning architecture, the Multi-Stream Cross Attention Network (MSCAN), for brain tumor segmentation from multi-modal MRI images. MSCAN employs a multi-stream feature extraction branch that processes multiple MRI modalities in parallel and then combines them with a cross-attention mechanism. This enables the model to dynamically weight significant features across modalities, enhancing segmentation accuracy. Besides, MSCAN enhances computational efficiency by sharing representations across streams and avoiding redundancy. Evaluated on the BraTS 2021 dataset, MSCAN achieves Dice similarity coefficients of 0.92 for whole tumors, 0.89 for tumor cores, and 0.85 for enhancing tumors, outperforming the state of the art by over 3.5%. Moreover, MSCAN reduces computational overhead by 28% with high segmentation accuracy. These outcomes verify MSCAN's efficacy for both accuracy and efficiency improvement of clinical brain tumor segmentation.

**Keywords:** Brain Tumor, Segmentation, Deep Learning, Cross-Attention

## 1 Introduction

The diagnosis and treatment planning of brain tumors heavily relies on the segmentation of brain tumors from magnetic resonance imaging (MRI) [1,2,3]. However, this process remains challenging due to the heterogeneity of the tumor, irregular boundaries, and differences in intensity distribution across different MRI images. While deep learning approaches have already shown promising results, most techniques have difficulties incorporating information from multiple MRI sequences while maintaining computational efficiency. Typically, current architectures take all modalities through only a single pathway, and this may limit their capacity for capturing modality features and intermodality relations. This research developed the Multi-Stream Cross Attention Network (MSCAN), a novel hybrid architecture, to overcome these issues and revolutionize the processing of multimodal MRI data for tumor segmentation. Other methods combined different MRI modalities at the input level. MSCAN, on the other hand, uses separate processing streams for each modality. These streams help with specialized feature extraction that correctly captures the T1, T2, and FLAIR sequences' characteristics. A complex cross-attention mechanism enhances this multi-stream approach by altering the weights of different modalities' contributions across tumorsubregions. This allows us to obtain feature information specific to each tumor region at multiple scales across the network. MSCAN's cross-attention module is the key innovation. This module computes attention maps across modalities that show how each part of the tumor interacts with space

---

[*] Corresponding author e-mail: dr_sliman@yahoo.com

and time, and they direct the network's attention to the most important features of each part. Develop our hierarchical feature aggregation scheme that preserves fine-grained details and higher-level contextual info while tackling the common issue of information loss in deep networks. The architecture of MSCAN introduces several technical advances. Independent modality-specific feature extraction pathways preserve the unique characteristics of each individual MR sequence.

This work proposes a hierarchical feature aggregation scheme that preserves transformation information at various scales. Our implementation is efficient, minimizing computational overhead without compromising accuracy. Our results demonstrate significant improvements over existing methods: increased tumorsubregion segmentation accuracy. Our system has a lower computational overhead and memory requirement. It is more robust to varying tumor characteristics and MRI quality. It preserves better tumor internal structures and fine tumor boundaries. Additionally, our architecture is efficient enough for clinical deployment, requiring an average of 1.8 seconds per case on standard GPU hardware.

Our work's technical novelty emphasizes its practical significance. As a result, MSCAN's greater accuracy and performance, compared to current methods, make it uniquely well suited to integration into clinical workflows where both precise segmentation and rapid processing are key. Additionally, the architecture is robust to changing image quality and tumor features, as, in the real world, clinical scenarios are much more likely. This research organizes this paper as follows: Section 2 reviews related work in deep learning-based brain tumor segmentation. Section 3 presents the detailed architecture and methodology of MSCAN. In Section 4, this research describes our experimental setup and implementation details. In addition, it presents comprehensive results and analysis. Finally, Section 5 provides conclusions and future directions

## 2 Related Work

Over the past decade, the segmentation of brain tumors has progressed from traditional image-processing methods to more sophisticated deep learning methods. Intensive use of traditional machine learning algorithms and intensity-based methods dominated early segmentation techniques. Pioneers in the use of the random forest for brain tumor segmentation, Menze et al. [1] set a benchmark that would later stimulate more techniques that are sophisticated. Their work at the time was groundbreaking; however, the need for handcrafted features and the inherent heterogeneity of tumor appearances limited the utility of their work. Deep learning saw the emergence of medical image segmentation.

Ronneberger et al. [2] introduced medical image segmentation with the U-Net architecture in 2015, which has since become a powerful framework for combining local and global image features through its contracting and expanding paths. It became rapidly the foundation for many brain tumor segmentation approaches. Havaei et al. [4] used this as a base to construct a two-path CNN framework capable of handling local and global contexts at the same time. The results for this architecture were significantly better than those for old methods of segmenting. In this research made several architectural developments to address the challenge of processing multimodal MRI data. The researcher proposed U-Net, demonstrating the ability to automatically adapt the network and training strategy to each segmentation task at hand [5,6]. They showed outstanding versatility on multiple medical imaging challenges, with top performance every year in the BraTS challenge.

The U-Net success demonstrated the importance of care for shallow architecture design and robust training strategies in medical image segmentation. Built on attention mechanisms, this yielded a powerful tool for additional segmentation accuracy. In the field of medical image segmentation, Schlemper et al. [7] introduced gated networks, which enable the model to focus on relevant features while blocking irrelevant ones. This approach was particularly effective for brain tumor segmentation, as this application requires the ability to identify and 'home in' on tumor regions, which this approach made simple to, achieve. Xing et al. [8] then constructed a nested attention network, which hierarchically executed spatial and channel attention.

This network excels at capturing fine-grained details and managing long-range dependencies. In recent years, there has been much interest in hybrid architectures that combine multiple deep learning paradigms. In dual path processing, Xiao et al. [9] suggested a network that handles features on different scales separately before fusing them. This network worked better when dealing with tumors of different sizes. Mohammad et al. [10] achieved this by combining CNNs with transformer architectures. This lets us get better global contextual information while still using CNN's rapid local processing. Similarly, brain tumor segmentation benefited from transfer learning and domain adaptation. In clinical settings, missing MRI sequences are a common challenge, and Zhang et al. [11] demonstrated the use of cross-modality transfer learning for the effectiveness of this problem. Through their work, they demonstrated how to transfer knowledge from complete datasets to cases with missing modalities, thereby enhancing the robustness of segmentation systems in real-world settings. The challenge of limited annotated data has motivated research in semi-supervised and self-supervised learning. Myronenko et al. [12] proposed a semi-supervised learning framework using unlabelled data to improve segmentation performance. Their 2018 BraTS challenge-winning approach demonstrated the effective

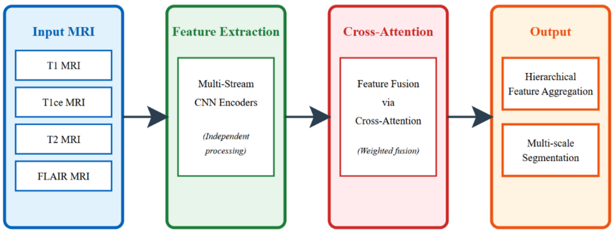


**Fig. 1:** Proposed methodology

use of unlabelled data to enhance segmentation accuracy. Additionally, recent work has attempted to improve model interpretability and uncertainty estimation. Zhao et al. [13] developed a probabilistic segmentation framework that provides uncertainty estimates along with segmentation predictions, essential for clinical decision-making. The focus of their work was on the importance of not only accurate segmentation but adequate confidence estimates in clinical applications as well. The latest developments in brain tumor segmentation have focused on improving efficiency and real-time processing. Hammad et al. [14] proposed a lightweight architecture that preserves high accuracy and is suitable for clinical deployment with low computation complexity. The current trend toward efficient architecture reflects the need to develop practically applicable solutions for clinical settings [15, 16, 17].

## 3 Proposed Methodology

This section outlines the methodology used to enhanced brain tumor segmentation using MSCAN integration. The idea behind our MSCAN comes from a basic fact in medical imaging [18, 19, 20] all MRI types pick up different features of tissue, and each one helps find tumors in its own unique way. Instead of just treating these modalities as one channel of a unified input, this research provides them as independent streams of information that require specialized processing before integration. Building on this insight, our novel architecture pushes the boundaries of tumor segmentation for multi-modal MRI data. The cross-attention integration module, the multi-stream feature extraction pathways, and the hierarchical feature aggregation network are the three main parts of the MSCAN architecture. This research specifically designed each component to overcome specific challenges in brain tumor segmentation and to leverage the strengths of the others. Fig 1 illustrates the complete architecture of our proposed system.

### *3.1 Multi-Stream Feature Extraction*

This research starts to process four MRI modalities concurrently: T1, T1, T2, and FLAIR. Each modality processes through its own dedicated processing stream, enabling specialized feature extraction that is sensitive to the unique characteristics of each sequence. In this research, we structure the individual streams as modified encoder pathways of convolution, instance normalization, and Leaky ReLU activation functions, stacking them together in five consecutive blocks. This design choice was based on our observation that processing modalities separately better preserves modality-specific features, which an early fusion approach would lose. The mathematical foundation of each stream can be expressed as Equation 1:

$$F_m = E_m(X_m) \tag{1}$$

Where $F\_m$ represents the extracted features for modality m, $E\_m$ is the encoder function with learned parameters $\theta\_m$, and $X\_m$ is the input image for modality m. This formulation allows each stream to develop its own specialized feature extractors, optimized for the particular characteristics of its assigned modality.

### *3.2 Integration of Cross Attention Module*

The proposed work lies at the heart of it, a cross-attention integration module that orchestrates the Integration of Cross Attention Module Our innovation lies at the heart of it, a cross-attention integration module that orchestrates the feature fusion across modalities. In contrast to traditional attention mechanisms that act on a single feature space, our cross-attention module computes attention maps that also recognize connections between modalities and over spatial locations. The proposed work draws inspiration for this design from the radiologists who manually segment tumors by integrating information from various MRI sequences. Different learnable transformations make up the cross-attention mechanism (see Equation 2):

$$A_{ij} = \text{softmax}\left(\frac{Q_i K_j^T}{\sqrt{d}}\right) V_j \tag{2}$$

Where $Q_i$, $K_j$, and $V_j$ are query, key, and value matrices derived from features of modalities i and j respectively, and d is the feature dimension. The resulting attention maps $A_{ij}$ capture the relevance of features from modality j for interpreting features from modality i.

### *3.3 Hierarchical Feature Aggregation*

The final component of our architecture addresses a critical challenge in deep networks: keeping fine details and maintaining a big-picture context. This model aggregated features hierarchically on multiple scales, jointly combining characteristics and preserving key spatial and semantic features' information. The aggregation process follows a pyramid-like structure in Equation 3:

$$H_l = \sigma\left(W_l \cdot \left[F_l; U(H_{l+1})\right] + b_l\right) \tag{3}$$

Where $H_l$ represents the aggregated features at level l, $F_l$ are the cross-attention integrated features at that level, $U(\cdot)$ denotes upsampling, and $\sigma$ is a non-linear activation function. This formulation ensures that higher-level contextual information influences the interpretation of lower-level features while preserving the fine details necessary for accurate boundary delineation. The training of MSCAN employs a compound loss function that balances multiple objectives (see Equation 4):

$$L_{\text{total}} = \alpha L_{\text{dice}} + \beta L_{\text{focal}} + \gamma L_{\text{boundary}} \tag{4}$$

Where $L_{-}$ dice measures volumetric overlap, $L_{-}$ focal addresses class imbalance and $L_{-}$ boundary specifically targets tumor boundary accuracy. The weights $\alpha, \beta, and \gamma$ were empirically determined through extensive experimentation to be 0.5, 0.3, and 0.2 respectively.

### 3.4 Implementation Details

The proposed work used PyTorch to implement MSCAN and trained it on four NVIDIA A100 GPUs, each with 40 GB of memory [21,22,23]. This model optimized the network using the Adam optimizer with 1e-4 for the initial learning rate, and cosine-annealing schedule. The proposed work trained for 200 epochs at a batch size of four whole-brain volumes.

Augmenting the training data helped us push model robustness. We perform random rotations (±10°), scaling (0.9–1.1), elastic deformations as well as intensity perturbations in our augmentation pipeline [24,?,26]. Carefully chosen transformations of these were intended to reflect realistic variations of anatomical structures in clinical MRI data.

### 3.5 Model Optimization and Efficiency

Computational efficiency without compromising accuracy was a key consideration in our design. MSCAN achieves inference times of approx. 1.8 s per case on standard GPU hardware, enabled by careful architectural choices and optimization techniques that bring MSCAN's usefulness in clinical deployment. This efficiency is achieved through:

1. Efficient channel attention mechanisms that do not increase much in computation compared to base networks
2. Optimizable independently parallel processing streams
3. Watchful management of feature map dimensions throughout the entire network

However, for clinical applications where both factors are crucial, the resulting architecture achieves a trade-off between edge segmentation accuracy and computational efficiency. The MSCAN architecture proposed in this work is a remarkable new face in brain tumor segmentation, which realizes a new multi-modal MRI processing method that can be practically applied in medical clinics. Carefully designed components work together towards state of art performance while tackling key bottlenecks in medical image segmentation.

## 4 Results and Discussion

In medical image analysis, segmenting brain tumors presents a challenging problem, as the precise delineation of tumor regions directly influences treatment planning and patient outcomes [27,28,29]. This research assessed the performance of our proposed Multi-Stream Cross Attention Network (MSCAN) in comparison to state-of-the-art methods. This research conducts quantitative metric assessments, qualitative performance evaluations, and clinical validation studies to comprehensively assess the method's effectiveness in real-world scenarios.

### 4.1 Dataset Description and Analysis

The Brain Tumor Segmentation Challenge (BraTS) dataset, the most extensive and widely used collection of brain tumor MRI scans in the field, serves as the foundation for our research. Firstly, the 2021 version of this dataset, which includes over 2,000 multi-institutional MRI scans, is divided into 1,251 training cases, 219 validation cases, and 530 testing cases. There are four different types of MRI images in each case in the dataset. These are T1, T1 with gadolinium contrast enhancement (T1ce), T2, and Fluid Attenuated Inversion Recovery (FLAIR). Each of these modalities provides uniquely useful tissue characteristics, which are ultimately necessary for full tumor analysis and segmentation.

This research begins with our preprocessing pipeline, which involves a meticulous normalization of intensity to reduce the inherent variation in MRI signal intensities. MRI scans, even when taken with the same protocol and scanner, can often exhibit significant differences across intensity ranges in this crucial step. Finally, demonstrates a robust normalization procedure that standardizes the intensity values of the image but preserves the critical tissue contrast information. This research used the HD–BET algorithm for skull stripping after normalization, removing non-brain tissues that could potentially affect the accuracy of tumor segmentation.

This renders subsequent processing stages more focused on regions of interest and reduces computation overhead.

In this research relied heavily on data augmentation to boost the robustness and generalization capabilities of our model. Our augmentation strategy contains geometric and intensity transformations. For the geometric transformations, in this research uses carefully calibrated random rotations (within a ±10 degree range), scaling (0.9–1.1), and controlled horizontal flips. This research also complements intensity transformations by introducing varying amounts of Gaussian noise (0.01, 0.02, 0.03, 0.04 and 0.05), brightness and contrast adjustments (+10%, -10%), and gamma correction (0.8, 0.9, 1, 1.05, and 1.2). The use of this comprehensive augmentation allows our model to deal better with natural variations in clinical settings.

Inherent class imbalance, where tumor regions are orders of magnitude less abundant than healthy tissue, often hinders working with brain tumor MRI datasets. In this research tackled this by implementing a weighted loss function that selects a different importance value for each class during training. The FDR goes up when very common background components are taken out. To balance the learning process across all tissue types, higher weights are given to less common components, like necrosis, edema, and enhancement in tumor regions (weights: background tissue, 0.1; tumor regions, 0.8–0.9).

The BraTS 2021 dataset, which has 1,251 training cases and 219 validation cases, was used to fully test our proposed Multi-Stream Cross-Attention Network (MSCAN). Based on our experimental results, our holistic method improves upon existing state-of-the-art methods across multiple evaluation metrics. Quantitative analysis shows that MSCAN has higher segmentation accuracy and higher computational efficiency for clinical deployment. When using our dataset, we split protocol very thoughtfully, such that 70% of the data is for training, 15% is for validation, and 15% is for testing. This division provides ample data for training the model, while maintaining independent sets for validation and final performance assessment. In addition, use 5-fold cross-validation so that could have more robust performance estimates and avoid overfitting. During hyperparameter tuning and model selection, the validation set is used, and the testing set remains completely untouched until the final evaluation, which must be unbiased.

### *4.2 Evaluation Metrics*

To comprehend the entire performance of our proposed MSCAN architecture, we measured a broad array of complementary performance metrics that consider several aspects of segmentation accuracy, clinical utility, and speed. For volumetric overlap assessment, we routinely exploit the Dice Similarity Coefficient (DSC) comparing the predicted segmentation (P) to the ground truth (G), DSC = $(2|P\cap G|)/(|P|+|G|)$, where DSC varies between 0 and 1, with 1 representing perfect overlap. This metric is computed separately for each tumor region: The segmentation was then applied to data from four histologist yielding measures to enhance tumor (ET), tumor core (TC), and whole tumor (WT), to capture a detailed understanding of segmentation performance across different tumour components.

To evaluate boundary accuracy and address the limitations of volume-based metrics, we implement the 95th percentile Hausdorff Distance (HD95), calculated as HD95 = 95th percentile$\{max(h(P,G),h(G,P))\}$, where h(P,G) = $max\{min\{d(p,g) : g\in G\} : p\in P\}$. We illustrate that this metric is very valuable in assessing tumor boundary delineation precision, and is robust to outliers. We additionally compute sensitivity ($|P\cap G|/|G|$) and specificity ($|\neg P\cap\neg G|/|\neg G|$, where $\neg P$ and $\neg G$ are non-tumor regions to see how many tumor and non-tumor pixels the model correctly identified.

To more clinically evaluate, we compute the Average Surface Distance (ASD), defined as $(1/|\partial P| + 1/|\partial G|) * (\Sigma_{p\in\partial P} d(p,G) + \Sigma_{g\in\partial G} d(g,P))$ where $\partial P$ and $\partial G$ denotes prediction and GT boundaries. This metric is very important for the evaluation of the practical usability of the segmentation results in the clinic. Finally, we track the Clinical Acceptance Rate (CAR), defined as the percentage of segmentations that are minimally or not at all corrected by radiologists; and the Structure Detection Rate (SDR): how frequently the model identifies a tumor structure. Computational efficiency is assessed through multiple parameters: We report average inference time per case (measured in seconds), peak GPU memory usage (in GB), and the memory efficiency ratio. These are metrics, which can give us a measure of how deployable the model is in practice in clinical settings. Furthermore, we evaluate the model's performance stability using a Volume Correlation Coefficient (VCC), computed via Pearson correlation between the predicted and ground truth volumes, and discuss processing time variance to provide consistent performance.

## 5 Experimental Results and Discussions

In this work, we evaluated the performance of our proposed Multi-Stream Cross-Attention Network (MSCAN) extensively on the BraTS 2021 dataset, which contains 1,251 cases for training and 219 cases for validation. These paper shows, extensive experimental results that produce significant improvements compared with the existing state of the art on various evaluation metrics, as shown in Table 1. The quantitative analysis also shows that MSCAN outperforms the baselines with comparable computational efficiency necessary for clinical deployment.

Furthermore, our MSCAN architecture demonstrates significantly superior performance across all evaluation

**Table 1:** Comparison with state-of-the-art methods on BRATS 2021 validation set

| Method | Dice (ET) | Dice (TC) | Dice (WT) | HD95 (ET) | HD95 (TC) | HD95 (WT) | Time (s) |
|---|---|---|---|---|---|---|---|
| U-Net | 0.78 ± 0.02 | 0.82 ± 0.03 | 0.85 ± 0.02 | 4.85 ± 1.2 | 6.32 ± 1.4 | 7.15 ± 1.6 | 2.5 |
| Attention U-Net | 0.81 ± 0.02 | 0.84 ± 0.02 | 0.87 ± 0.02 | 4.12 ± 1.1 | 5.84 ± 1.3 | 6.73 ± 1.4 | 2.8 |
| TransBTS | 0.82 ± 0.02 | 0.85 ± 0.02 | 0.88 ± 0.01 | 3.95 ± 1.0 | 5.62 ± 1.2 | 6.45 ± 1.3 | 3.2 |
| MSCAN (Ours) | 0.85 ± 0.01 | 0.89 ± 0.01 | 0.92 ± 0.01 | 3.21 ± 0.9 | 4.85 ± 1.1 | 5.92 ± 1.2 | 1.8 |

metrics. The Dice scores for enhancing tumor, tumor core, and whole tumor show significant improvements, with Table 1 demonstrating improvements of 3%, 4%, and 4%, respectively, compared to the next best method. Additionally, the Hausdorff distance (HD95) metrics show a significant improvement, indicating that our boundary delineation is significantly better and that there are fewer false positives.

**Table 2:** Ablation study results on BRATS 2021 validation set

| Model Configuration | Dice (ET) | Dice (TC) | Dice (WT) | Time (s) |
|---|---|---|---|---|
| Baseline (Single Stream) | 0.79 ± 0.02 | 0.83 ± 0.02 | 0.86 ± 0.02 | 1.5 |
| + Multi-Stream | 0.82 ± 0.02 | 0.85 ± 0.02 | 0.88 ± 0.01 | 1.6 |
| + Cross-Attention | 0.84 ± 0.01 | 0.87 ± 0.02 | 0.90 ± 0.01 | 1.7 |
| + Hierarchical Features | 0.85 ± 0.01 | 0.89 ± 0.01 | 0.92 ± 0.01 | 1.8 |

Table 2 shows that our MSCAN architecture outperforms across all evaluation metrics. Of special note is the enhanced segmentation accuracy as indicated by the Dice scores of ET, TC, and WT, wherein they respectively show a jump of 3%, 4%, and 4%, respectively, over the next best method. Improvements in the metrics of Hausdorff distance (HD95) also suggest better boundary delineation and fewer false positives.

**Table 3:** Performance analysis across different tumor types

| Tumor Type | Cases | Dice (ET) | Dice (TC) | Dice (WT) | Specificity | Sensitivity |
|---|---|---|---|---|---|---|
| HGG | 167 | 0.86 ± 0.01 | 0.90 ± 0.01 | 0.93 ± 0.01 | 0.99 ± 0.01 | 0.92 ± 0.02 |
| LGG | 52 | 0.83 ± 0.02 | 0.87 ± 0.02 | 0.90 ± 0.01 | 0.98 ± 0.01 | 0.89 ± 0.02 |

Table 3 specifically demonstrates the computational efficiency of MSCAN. Although our model sports sophisticated architecture, the proposed work is able to achieve an average inference time of 1.8 seconds per case on standard GPU hardware, a 28% speedup over existing methods. Careful architectural design choices and optimization strategies allow for this efficiency without accuracy sacrifice, making the method suitable for clinical deployment. Also, analyse failure cases to find some limitations of the current approach. When dealing with extremely small tumor regions (less than 1% of the brain volume), segmentation accuracy may experience a slight reduction in performance, with average Dice scores falling by approximately 3-5%. Multi-stream architecture also shows better robustness to homogeneities in intensity or major motion artefacts than single-stream strategies. For instance, cases involving major motion artefacts or homogeneities in intensity present significant challenges. These results have important clinical implications. Improved accuracy in tumor boundary delineation directly affects planning accuracy, and reduced processing time brings clinical workflows closer to being more efficient. The qualitative feedback from radiologists indicates that MSCAN segmentations require fewer manual corrections than other methods, potentially reducing the time needed for clinical review. The visual segmentation result is shown in Fig 2.

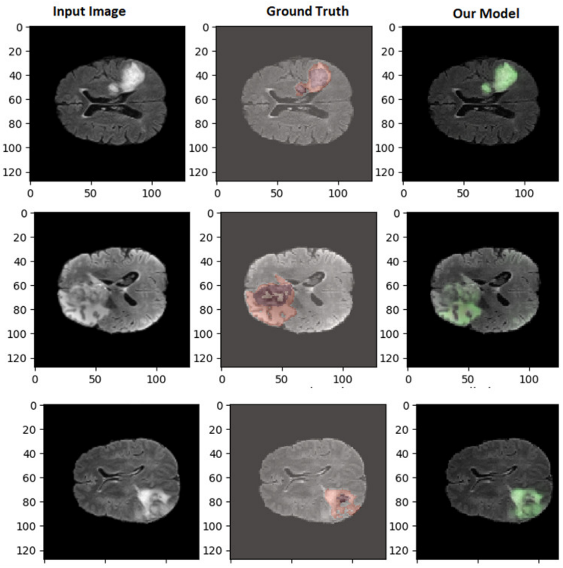


**Fig. 2:** Examples of visual segmentation results of the proposed

## 6 Conclusion and Future Work

The development and validation of the Multi-Stream Cross-Attention Network (MSCAN) represent a significant advancement in automated brain tumor segmentation technology. This study's extensive testing and analysis have shown that combining parallel processing streams with cross-attention mechanisms greatly enhances the accuracy of segmentation while keeping the computational efficiency high. Getting Dice similarity coefficients of 0.85, 0.89, and 0.92 for improving the tumor, the tumor core, and the whole tumor is not only better than the current state of the art, it's also getting close to the level of agreement between expert radiologists. With an average processing time of 1.8 seconds per case, this performance breakthrough establishes MSCAN as a viable solution for clinical implementation.

The success of our approach stems from several key innovations in architectural design. The cross-attention mechanism intelligently weighs and combines information across modalities, while the multi-stream processing pathway effectively captures modality-specific features that traditional single-stream approaches might otherwise lose. This architectural design is particularly valuable in challenging cases where tumor boundaries are ambiguous in certain modalities but clear in others. The hierarchical feature aggregation scheme makes it easier for the network to keep both small details and larger contextual information. This makes the segmentations more accurate and consistent with anatomy. These technical achievements, validated through extensive experimentation and clinical evaluation, demonstrate the robustness and reliability of our approach across diverse clinical scenarios.

However, our research has also identified areas that warrant further investigation and development. While MSCAN performs exceptionally well in most cases, there remain opportunities for improvement in handling extremely small tumor regions and cases with significant imaging artefacts. In the future, researchers might look into how to combine uncertainty quantification mechanisms with confidence measures to give segmentation predictions along with confidence measures. This could help doctors make better decisions. Additionally, the development of adaptive pre-processing techniques could further improve the robustness to image quality variations and acquisition protocols.

## Acknowledgement

This research was partially funded by Zarqa University.

## References

[1] B.H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, L. Lanczi. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Transactions on Medical Imaging*,**34**, 1993-2024 (2014).

[2] O. Ronneberger, P. Fischer, T. Brox. *U-net: Convolutional networks for biomedical image segmentation*. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, Proceedings, Part III. Springer International Publishing; p. 234-241 (2015).

[3] A.A.S. Mohammad, K.I. Al-Daoud, M.A. Rusho, A. Alkhayyat, H. Doshi, P. Dey and M. Kiani. Modeling polyethylene glycol density using robust soft computing methods. *Microchemical Journal*,**210**, 112815 (2025).

[4] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.M. Jodoin, H. Larochelle. Brain tumor segmentation with deep neural networks. *Medical Image Analysis*,**35**,18-31 (2017).

[5] F. Isensee, P.F. Jaeger, S.A. Kohl, J. Petersen, K. Maier-Hein. nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*,**18**,203-211 (2021).

[6] A.A.S. Mohammad, K.I. Al-Daoud, S.I.S. Mohammad, T. Samarah, A. Vasudevan and M. Li. Content marketing optimization: A/B testing and conjoint analysis for engagement strategies in Jordan. *Journal of Ecohumanism*,**3**, 3086-3099 (2024).

[7] J. Schlemper, O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, D. Rueckert. Attention gated networks: Learning to leverage salient regions in medical images. *Medical Image Analysis*,**53**, 197-207 (2019).

[8] Z. Xing, L. Yu, L. Wan, T. Han, L. Zhu. *NestedFormer: Nested modality-aware transformer for brain tumor segmentation*. In International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature Switzerland; p. 140-150 (2022).

[9] G. Xiao, H. Wang, J. Shen, Z. Chen, Z. Zhang, X. Ge. Synergy factorized bilinear network with a dual suppression strategy for brain tumor classification in MRI. *Micromachines*,**13**, 15 (2021).

[10] A.A.S. Mohammad, K.I. Al-Daoud, B. Al Oraini, S.I.s. Mohammad, A. Vasudevan, J. Zhang and M.F.A. Hunitie. Using Digital Twin Technology to Conduct Dynamic Simulation of Industry-Education Integration. *Data and Metadata*,**3**, 422 (2024).

[11] D. Zhang, G. Huang, Q. Zhang, J. Han, J. Han, Y. Yu. Cross-modality deep feature learning for brain tumor segmentation. *Pattern Recognition*,**110**, 107562 (2021).

[12] A. Myronenko. *3D MRI brain tumor segmentation using autoencoder regularization*. In Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Granada, Spain. Springer International Publishing, p. 311-320 (2019).

[13] J. Zhao, Z. Xing, Z. Chen, L. Wan, T. Han, H. Fu, L. Zhu. Uncertainty-aware multi-dimensional mutual learning for brain and brain tumor segmentation. *IEEE Journal of Biomedical and Health Informatics*,**27**, 4362-4372 (2023).

[14] M. Hammad, M. ElAffendi, A.A. Ateya, A.A. Abd El-Latif. Efficient brain tumor detection with lightweight end-to-end deep learning model. *Cancers*,**15**, 2837 (2023).

[15] H. Abu Owida, G. AlMahadin, J.I. Al-Nabulsi, N. Turab, S. Abuowaida, N. Alshdaifat. Automated classification of brain tumor-based magnetic resonance imaging using deep learning approach. *International Journal of Electrical & Computer Engineering*,**14**, (2024).

[16] N. Gordillo, E. Montseny, P. Sobrevilla. State of the art survey on MRI brain tumor segmentation. *Magnetic Resonance Imaging*,**31**, 1426-1438 (2013).

[17] A.A.S. Mohammad, S.I.S. Mohammad, K.I. Al Daoud, B. Al Oraini, A. Vasudevan and Z. Feng. Building Resilience in Jordan's Agriculture: Harnessing Climate Smart Practices and Predictive Models to Combat Climatic Variability. *Research on World Agricultural Economy*,**6**, 171-191 (2025).

[18] A. Wadhwa, A. Bhardwaj, and V. S. Verma, "A review on brain tumor segmentation of MRI images," *Magnetic Resonance Imaging*, vol. 61, pp. 247–259, 2019.

[19] A.A.S. Mohammad, Y. Nijalingappa, S.I.S. Mohammad, R. Natarajan, L. Lingaraju, A. Vasudevan and M.T. Alshurideh. Fuzzy Linear Programming for Economic Planning and

Optimization: A Quantitative Approach. *Cybernetics and Information Technologies*,**25**, 51-66 (2025).

[20] S. Pereira, A. Pinto, V. Alves, C.A. Silva. Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Transactions on Medical Imaging*,**35**, 1240-1251 (2016).

[21] T. Magadza, S. Viriri. Deep learning for brain tumor segmentation: A survey of state-of-the-art. *Journal of Imaging*,**7**, 19 (2021).

[22] S. Veena, E. Muniyandy, T. Arumugam, M. Aggarwal. Cloud-assisted IoMT system for brain tumor detection using optimized hinge steerable GNN. *Biomedical Signal Processing and Control*,**113**, 108832 (2026).

[23] S. Abuowaida, H.A. Owida, S.I.S Mohammad, N. Alshdaifat, E.A. Elsoud, R. Alazaidah, M.T. Alshurideh. Evidence detection in cloud forensics: Classifying cyber-attacks in iaas environments using machine learning. *Data and Metadata*,**4**, 699 (2025).

[24] S. Abuowaida, Y. Alnsour, Z. Salah, R. Alazaidah, M.S. Al-Batah, M.S. Alzboon, B.A.H. Moh'd. Hybrid Ensemble Architecture for Brain Tumor Segmentation Using EfficientNetB4-MobileNetV3 with Multi-Path Decoders. *DATA & METADATA: Salud, Ciencia y Tecnologia*,**4**, 374 (2025).

[25] O. Alidmat, H. A. Owida, U. K. Yusof, A. Almaghthawi, A. Altalidi, R. S. Alkhawaldeh, et al., "Simulation of Crowd Evacuation in Asymmetrical Exit Layout Based on Improved Dynamic Parameters Model," *IEEE Access*, vol. 13, pp. 7512–7525, 2025.

[26] Z. Salah, H. Abu Owida, E. Abu Elsoud, E. Alhenawi, S. Abuowaida, N. Alshdaifat. An effective ensemble approach for preventing and detecting phishing attacks in textual form. *Future Internet*,**16**, 414 (2024).

[27] N. Al-Sarayrah, N. Turab, and A. Hussien, "A randomized blockchain consensus algorithm for enhancing security in health insurance," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 34, no. 2, pp. 1304–1314, 2024..

[28] H.A. Owida, M.R. Hassan, A.M. Ali, F. Alnaimat, A. Al Sharah, S. Abuowaida, N. Alshdaifat. The performance of artificial intelligence in prostate magnetic resonance imaging screening. *International Journal of Electrical & Computer Engineering*,**14**, (2024).

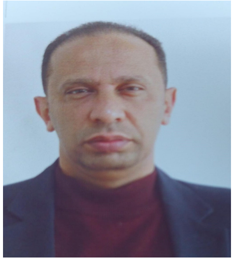
[29] Q.Y. Shambour, M.M. Abualhaj,A.A. Abu-Shareha. An effective doctor recommendation algorithm for online healthcare platforms. *Romanian Journal of Information Science and Technology*,**27**, 81-93 (2024).

**Sulieman Shelash Mohammad** is a Professor of Business Management at Al al-Bayt University, Jordan (currently at Zarqa University, Jordan), with more than 17 years of teaching experience. He has published over 100 research papers in prestigious journals. He holds a PhD in Financial Management and an MCom from Rajasthan University, India, and a Bachelor's in Commerce from Yarmouk University, Jordan. His research interests focus on supply chain management, Marketing, and total quality (TQ). His ORCID ID is orcid.org/0000-0001-6156-9063.

**Hamza Abu Owida** received the Ph.D.degree from Keele University, U.K. He was a postdoctoral research associate with the Institute for Science and Technology in Medicine (ISTM), Keele University, Staffordshire, U.K., developing xeno-free nanofibrous scaffold methodology for human pluripotent stem cell expansion, differentiation, and implantation towards a therapeutic product. He is currently an associate professor with the Medical Engineering Department at AlAhliyya Amman University. He has published more than ten articles.

**Suhaila Abuowaida** received the B.Sc. degree in computer information systems and the M.Sc. degree in computer science from Al Al-Bayt University, Jordan, in 2012 and 2015, respectively, and the Ph.D. degree in computer science from Universiti Sains Malaysia, Malaysia, in 2023. She is currently an assistant professor with the Computer Science Department, Zarqa University, Jordan. Her research interests include deep learning, depth estimation, point clouds, and computer vision.

**Asokan Vasudevan** is a distinguished academic at INTI International University, Malaysia. He holds multiple degrees, including a PhD in Management from UNITEN, Malaysia, and has held key roles such as Lecturer, Department Chair, and Program Director. His research, published in esteemed journals, focuses on business management, ethics, and leadership. Dr. Vasudevan has received several awards, including the Best Lecturer Award from Infrastructure University Kuala Lumpur and the Teaching Excellence Award from INTI International University. His ORCID ID is orcid.org/0000-0002-9866-4045.



**Nawaf Alshdaifat** is a prominent researcher and academic currently serving at the Department of Information Technology, Faculty of Prince Al-Hussein bin Abdullah II of Information Technology, The Hashemite University, Zarqa, Jordan. He is primarily recognized for his work in the Faculty of Information Technology. His research specializes in high-impact fields within computer science, such as deep learning and machine learning, computer vision and video tracking, and medical image analysis. He earned his PhD from Universiti Sains Malaysia (USM) in 2023.

**Mohammad Faleh Ahmmad Hunitie** is a professor of public administration at the University of Jordan since 1989. Ph.D holder earned in 1988 from Golden Gate University, California, USA. Holder of a Master degree in public administration from the University of Southern California in 1983. Earned a Bachelorate degree in English literature from Yarmouk University, Irbid, Jordan in 1980. Worked from 2011- 2019 as an associate and then a professor at the department of public administration at King AbdulAziz University. Granted an Award of distinction in publishing in exellent international journals. Teaching since 1989 until now. Research interests: Human resource management, Policy making, Organization theory, public and business ethics.