Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 21 No. 5 (2025) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v21i05.53405

PAPER

Brain Tumour Segmentation and Edge Detection Using Self-Supervised Learning

Dasith Samarasinghe[1], Deshan Wickramasinghe[1], Theshan Wijerathne[1], Dulani Meedeniya[1], Pratheepan Yogarajah[2](✉)

[1]Department of Computer Science and Engineering, University of Moratuwa, Moratuwa, Sri Lanka

[2]School of Computing, Engineering and Intelligent Systems, Ulster University, Londonderry, United Kingdom

p.yogarajah@ulster.ac.uk

ABSTRACT

Brain tumour segmentation is critical in medical image analysis, facilitating diagnosis and treatment planning in neurosurgery. Brain tumour segmentation with supervised learning shows robust results in medical imaging; however, it requires a sufficient amount of annotated data for effective learning. It is important to detect boundaries of tumour subregions accurately in fine-grained segmentation. We propose a novel approach that uses a unique dual-decoder architecture, focusing on edge identification and segmentation accuracy enhancement. Utilising a dual-decoder 3D-UNet model, we prioritise accuracy and fine-grained details in tumour segmentation and introduce an additional tumour edge detection task, aiming to move beyond traditional single-decoder approaches. Incorporating a 3D SimSiam network as the self-supervised pretraining technique, we aim to address the limitation of annotated data and enhance the segmentation accuracy. Our model surpasses many supervised variants of U-net architectures and self-supervised approaches, highlighting the importance of edge detection in tumour segmentation. The proposed approach enhances segmentation accuracy by showing an accuracy of 98.1% and provides critical boundary details for clinical decision-making. Visualisations of segmentation and edge masks further validate the effectiveness of the proposed method.

KEYWORDS

artificial intelligence (AI), contrastive learning, deep learning, dual-decoder, multi-modal, magnetic resonance imaging (MRI)

1 INTRODUCTION

Brain tumours are unusual growths of cells that can disrupt vital brain structures. These can be classified as either high-grade gliomas (HGG) or low-grade gliomas (LGG) [1]. Locating the brain tumours is crucial because as they grow, they might press against key brain structures. However, it has become challenging to identify brain tumours due to their location-specific nature. However, brain tumour segmentation using 3D magnetic resonance imaging (MRI) has played a significant role in

Samarasinghe, D., Wickramasinghe, D., Wijerathne, T., Meedeniya, D., Yogarajah, P. (2025). Brain Tumour Segmentation and Edge Detection Using Self-Supervised Learning. *International Journal of Online and Biomedical Engineering (iJOE)*, 21(5), pp. 127–141. https://doi.org/10.3991/ijoe.v21i05.53405

Article submitted 2024-11-19. Revision uploaded 2025-02-07. Final acceptance 2025-02-07.

addressing this issue [2]. Segmentation of brain tumours using medical images has allowed clinicians to accurately locate tumour boundaries and carry out treatment plans [3]. Many studies have been conducted on brain tumour segmentation using a 3D convolution neural network (CNN) with limited annotated 3D MRI data. Due to the high cost of manual labelling and privacy concerns, there are a limited number of annotated MRIs. To address these limitations, self-supervised learning (SSL) emerges as a promising approach, leveraging unlabelled medical images for pretraining and mitigating dependency on annotated data. Common SSL techniques include pretext-based tasks, contrastive learning, and redundancy reduction [4], each offering distinct advantages. There is a gap between the pretext task and the segmentation downstream task, which makes it challenging to ensure that the learnt representations are directly relevant and beneficial for the downstream task. Most of the contrastive learning approaches require the generation of positive and negative pairs [5].

In this study, we leverage the SimSiam model, a contrastive learning approach, for pretraining a deep learning model on unlabelled data. The exploration of the literature shows that determining the tumour boundaries is crucial for accurate tumour segmentation. Our solution hinges on a sophisticated encoder-decoder architecture, which serves as its core. We incorporate two distinct decoders: "edge decoder" and "segmentation decoder". The edge decoder is specifically designed for extracting edge masks, whereas the segmentation decoder focuses on learning significant features to improve segmentation accuracy. This innovative architecture aims to optimise boundary identification and improve overall segmentation performance. This paper proposes an approach integrating SSL with a specialised encoder-decoder architecture for robust and accurate brain tumour segmentation, demonstrating its potential for enhancing clinical practices and patient outcomes.

2 BACKGROUND AND RELATED STUDIES

The use of multi-modal 3D images is crucial for accurately capturing the intricate structures of the brain. This approach is especially effective in diagnosing serious brain tumours, which require precise analysis. However, the large amount of MRI data processing is challenging in practice. Manually labelling this data, which involves defining specific regions for computer-aided systems, is time-consuming. Current research in artificial intelligence (AI) for medical applications heavily depends on analysing labelled MRI data provided by clinicians. This process is foundational but limited by the availability and accuracy of these labels. To address this, implementing a model that utilises multi-modal image segmentation on unlabelled data can significantly enhance the accuracy and usability of brain tumour diagnosis systems.

Moreover, healthcare professionals need AI systems that offer clear and understandable reasoning behind their decisions. This explainability is essential for several reasons. It helps validate the AI's recommendations, ensures patient safety, and builds trust among medical professionals. An explainable AI model provides the transparency needed to integrate AI technologies into clinical practice confidently. By leveraging multi-modal 3D imaging and developing SSL models, we can create a more effective and reliable tool for brain tumour segmentation. This advancement not only promises to improve diagnostic accuracy but also aligns with the growing need for transparent and trustworthy AI in healthcare.

A crucial aspect often overlooked in brain tumour segmentation applications is the inclusion of tumour edge images as an output. Most existing applications focus

solely on providing segmented tumour images, which offer a broad view of the affected area but lack the precise delineation of tumour boundaries. Edge detection images, however, provide a clear and accurate contour of the tumour, which is vital for several reasons. Firstly, precise edge information is essential for surgical planning and radiotherapy, as it helps in minimising the risk of damaging healthy brain tissue during treatment. Secondly, edge images significantly enhance the interpretability of the segmentation results. They offer additional clarity by highlighting the specific contours of the tumour, which supports better decision-making and diagnostic accuracy. The inclusion of tumour edge images aligns with the principles of explainable AI, offering medical professionals deeper insights into the model's outputs. This additional layer of information is crucial for effective and safer patient care, and its absence in many applications underscores the need for advancements in this area. To build a strong foundation for our approach, it is essential to explore existing related literature, highlighting key developments, successful techniques, and ongoing challenges in brain tumour segmentation and explainable AI. Table 1 summarises related studies on brain tumour segmentation.

Table 1. Summary of aelated studies

Study	Dataset	Techniques	Performance
[4]	BraTS 2018	3D Relative Patch Location	DC: 81.28% (Tumor Core) DC: 86.12% (Enhanced)
[5]	BraTS 2020	Graph NN, CNN, Transformer	DC: 92.01%
[6]	BraTS 2021	Swin UNETR	DC: 90.36%
[7]	BraTS 2021, MSD 2019	MOD (tokenizer & dense predictor)	DC: 92.62% DC: 80.36%
[8]	BraTS 2018/2019/2020	K-means, Fusion with Naïve Bayesian	DC: 91.25%
[9]	BraTS 2018	BT-UNet	DC: 70.1%
[10]	BraTS 2020, Kaggle LGG, BTC_postop	Hybrid Attention-Residual U-NET with Transformer Blocks	DC: 96% DC: 97% DC: 88%
[11]	Kaggle LGG,	ResNet50 + classification header	Accuracy: 94% Precision: 92%
[12]	CE-MRI	Multi-path CNN	Accuracy: 97%

The common SSL approach involves pre-training the encoder part, followed by training the full network with a limited amount of labelled data for the downstream task. There are three major types of SSL-based strategies, namely (1) pretext-based tasks, (2) contrastive learning, and (3) redundancy reduction. Taleb et al. [4] introduced a pretext-based task as Relative 3D Patch Location (RPL). 3D RPL generates non-overlapping 3D random patches. Leveraging a centre reference patch, the task involves classifying the spatial relationships between patches and treating the location of other patches as positive labels against all other patch locations as negatives. For the tumour core (TC) and the enhancing tumour (ET) regions, 3D-RPL achieved DICE scores of 81.28% and 86.12%, respectively.

In another study, Punn et al., [9] have introduced the BT-UNet, where the Barlow Twins approach is coupled with U-Net segmentation models for effective medical image segmentation. Using a limited number of annotated data samples, the BT-Unet framework fine-tunes the U-Net model for downstream segmentation of biomedical

image tasks. This is achieved by pre-training the encoder of the U-Net with features of the data in an unsupervised manner, based on the previously described redundancy reduction-based Barlow Twins strategy. The pre-training step learns complicated feature representations from unannotated data sets. This phase focuses on training the U-Net model's encoder network using the Barlow Twins SSL technique. The specific results of this study on brain tumour segmentation claim that to achieve better results, the encoder part should be complex enough, which encourages the use of a complex U-net type. They have shown this by applying the BT concept to U-net types of different complexities and performing downstream segmentations.

However, pretext-based methods face difficulties due to the mismatch between pre-training tasks and downstream objectives, limiting their efficacy. Contrastive learning offers great adaptability by fine-tuning representations to suit different downstream tasks, enhancing versatility. On the other hand, redundancy reduction techniques prioritise feature redundancy minimisation over task-specific representation learning, potentially constraining their utility in complex scenarios.

3 METHODOLOGY

3.1 Datasets

The brain tumour segmentation (BraTS) [13] dataset of the Multimodal Brain Tumour Segmentation Challenge is a widely utilised benchmark dataset in medical image domain analysis, specifically focusing on brain tumour segmentation in MRI scans. We utilised BraTS 2018 and BraTS 2020 publicly available datasets. The BraTS 2018 dataset consists of 285 training images and 66 validation images, while the BraTS 2020 dataset contains 369 training images and 125 validation images. Each dataset consists of multi-modal MRI scans, including T1, T2, T1ce (T1 with gadolinium contrast enhancement), and Flair images. The segmentation labels provided for evaluation were categorised into four classes: background and three tumour subregions, WT, TC, and enhancing tumour.

3.2 Data preprocessing

Firstly, all the original images were centre-cropped to the dimension of 128 × 128 × 128 to exclude irrelevant background regions, reducing the computational burden during training and focusing the model on the brain parenchyma, where tumours are most likely to reside. Then we performed normalising each image of the dataset. Normalization is essential for standardising the intensity values across different medical images [14]. In our pipeline, we employed min-max scaling to map the intensity values in each modality to a specific range (between 0 and 1). This ensures that all images have comparable intensity distributions, facilitating the training process and preventing instabilities within the models. We also separated the original training dataset into train and validation sets with ratios of 75 and 25, respectively.

3.3 Edge mask label generation

Following the novel two-decoder approach we have suggested, labelled edge mask data were required while training the model at the downstream phase. However, the original BraTS datasets contain only the labelled segmentation masks,

and therefore a separate algorithm was applied to generate corresponding edge masks for each segmentation mask. We used a 3D Laplacian filter-based algorithm for the edge mask generation process as proposed in [15], where each segmentation mask was convolved with the special 3D Laplacian-like filter shown in Figure 1 to generate the corresponding 3D edge masks. The resulting edge masks were then one-hot encoded into 4 target classes: background, WT, TC, and enhancing tumour.

$$\left[\begin{bmatrix}-1 & -1 & -1\\ -1 & -1 & -1\\ -1 & -1 & -1\end{bmatrix}\begin{bmatrix}-1 & -1 & -1\\ -1 & 26 & -1\\ -1 & -1 & -1\end{bmatrix}\begin{bmatrix}-1 & -1 & -1\\ -1 & -1 & -1\\ -1 & -1 & -1\end{bmatrix}\right]$$

Fig. 1. Laplacian filter to generate the 3D edge mask

3.4 Self-supervised pretraining – simsiam network

We employed SSL to pre-train the model on unlabelled medical images before performing the actual segmentation task. This approach aims to learn important representations of brain anatomy from the unlabelled data, which can then improve the performance of the segmentation model on limited labelled data.

We modified the encoder architecture of the original SimSiam model to use it with 3D MRI images. The encoder aims to learn a low-dimensional embedding that captures the essential features of the input image. Therefore, we created two datasets from the initial data that are two different augmented versions of each other acting as positive pairs. Since the BraTS2020 dataset consists of 369 images, two dataset versions double the total data count we used for this process. In our augmentation pipeline, we employed two data augmentation techniques, random rotation and random zooming. The compressed representation of the encoder is then fetched into a predictor network to predict the embedding of the other image in the pair. The key metric to the training process is the contrastive loss, which is negative cosine similarity as mentioned by the authors of the SimSiam model. Our encoder was trained to optimise the negative cosine similarity between our dataset's two distinct versions.

We utilised 3D MRI data of all 4 modalities of the BraTS 2020 dataset without segmentation labels since the pretraining phase entirely depends on unlabelled data. Hyperparameter tuning in the context of a 3D SimSiam network involves optimising various parameters that govern the training process to enhance model performance. We trained the model with different initial learning rates (LRs) such as 0.0001, 0.001, and 0.002 for 300 epochs., focusing on the factor that longer training times are beneficial for SimSiam pretraining. Here, the LR is crucial for controlling the change in the model in response to the estimated error each time the model weights are updated, aiming for fast convergence. Tuning epochs is essential to prevent overfitting or underfitting. Further, Bayesian optimisation is used.

3.5 Segmentation – Dual-decoder 3D-UNet

The encoder-decoder structure is a widely used deep learning architecture for semantic segmentation. It consists of a feature extraction phase (encoder) and a feature reconstruction phase (decoder) [16]. The encoder-decoder structure can be applied to both 2D and 3D images and can be combined with various modules such as skip connections, dilated convolutions, attention mechanisms, and dense connections to improve segmentation performance.

The 3D-UNet [16] architecture is a widely used encoder-decoder architecture for semantic segmentation tasks, particularly in the field of biomedical image analysis. In contrast to typical U-net models, in our model architecture, we have two separate decoders, referred to as the "segmentation decoder" and "edge decoder". Edge decoder specialises in extracting edge masks, while segmentation decoder is dedicated to learning meaningful features that enhance segmentation accuracy.

The input data is in four modalities with the dimensions of 128 × 128 × 128 × 4. There are two different types of labels to be provided for two decoder branches. The segmentation decoder will be provided with segmentation mask labels, and the edge decoder will be fetched with edge mask labels. The model utilises a single encoder shared with both decoders as shown in Figure 2. The encoder has been pre-trained in a self-supervised manner as mentioned in the above section, and the pre-trained encoder weights of the SimSiam model were transferred to related layers. The encoder consists of five convolutional blocks with multiple 3D convolution layers, rectified linear unit (ReLU) activation, and dropout for regularisation. The max pooling layers downsample the feature maps after each block, capturing context at different scales. Therefore, the shared encoder extracts features from the input data and then shares them with both decoders. Although ground truth segmentation labels of the dataset have four distinct classes as mentioned in Section 3, the background was the most prominent class due to the limited region acquired by the tumour. Therefore, we utilised class weights while calculating the dice loss to handle the class imbalance, focusing more on tumour subregions.

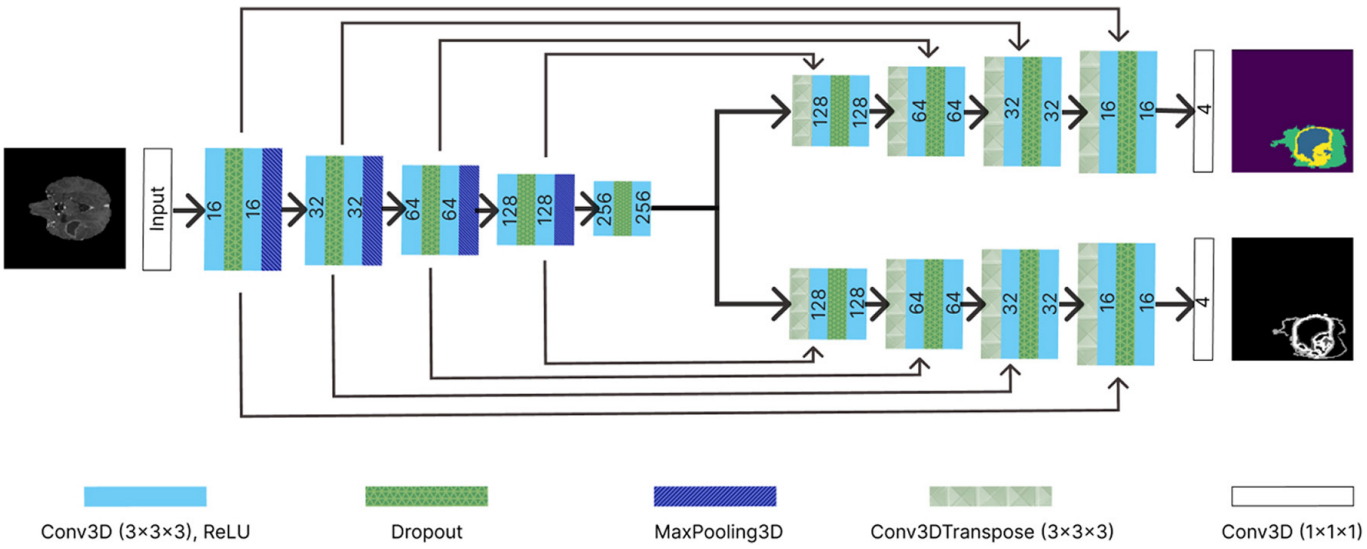

Fig. 2. High-level architecture of the proposed dual-decoder 3D U-net model

As shown in Figure 2, the segmentation decoder specialises in segmentation, employing four upsampling blocks with Conv3D. Transpose layers, skip connections to recover spatial details, and culminate in a 1 × 1 × 1 convolution with softmax activation to produce the final segmentation mask. As the labels, we used segmentation masks that have the same dimension (128 × 128 × 128 × 4) as the input data. The Simsiam pre-trained weights support potential segmentation improvement.

In order to effectively handle the dominance of the background class in the model, class weights are calculated using inverse frequency and applied dice loss to adjust the influence of each class during training. Given four classes of background (0), necrotic and non-enhancing tumour core (1), oedema (2), and enhancing tumour (3), we calculate weights as Wi = N/(C × Ni), where N is the total pixel count, C is the

number of classes, and Ni is the pixel count of a given class. This ensures that infrequent tumour regions are assigned higher weights while the dominant background class is assigned a relatively lower weight. Furthermore, we normalised these weights between 0 and 1 for stability and consistency of the training process. The computed class weights are then integrated into the Dice Loss function, a major loss function used in our combined total loss function. Weighted Dice Loss ensures that infrequent tumour classes contribute more significantly to the loss, improving segmentation accuracy. Thus, we effectively reduce the dominance of the background class and improve segmentation performance across all tumour regions.

On the other hand, the edge decoder has a similar structure to the segmentation decoder, with upsampling and skip connections, and the final $1 \times 1 \times 1$ convolution with softmax activation for the edge mask. The overall approach shown in Figure 3 illustrates how the self-supervised pretraining combines with the downstream segmentation and edge detection process, where the meaningful representations learnt by the SimSiam model have been transformed into the dual-decoder U-net model to improve the segmentation process using a lesser amount of labelled data.

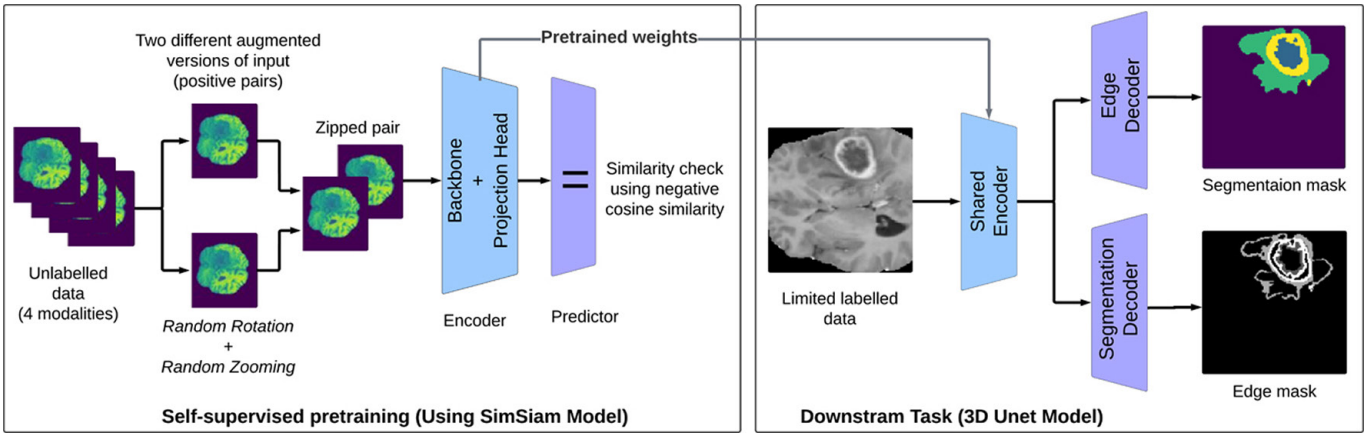

Fig. 3. The overall approach of how the self-supervised pretraining combines with the downstream segmentation and edge detection process

4 RESULTS AND DISCUSSION

4.1 SimSiam pretraining results

During the pretraining, we observed a phenomenon in this SimSiam model. For relatively higher initial LRs, the model loss, which is negative cosine similarity, reached –1 very quickly in less than 25 epochs. The reason for that was the encoder yielded almost identical outputs for all the input images, which is known as the "representation collapse" of the encoder. This will result in our encoder learning very few or no new structural representations at all, which will fail the pretraining process. We found out that relatively lower LRs can help to avoid the representation collapse of the encoder. We pretrained the model for 300 epochs with an initial LR of 0.0001, utilising the BraTS 2020 dataset, and achieved a cosine similarity of –0.995.

4.2 Dual-decoder 3D-UNet results

Implementation of the downstream task was done using the Keras API with the TensorFlow backend. The LR was set to 0.0001 and trained for 100 epochs, with a

batch size of 1 due to resource constraints. All the models were trained on an NVIDIA RTX 2080 Super GPU with 6GB VRAM.

We experimented with our dual-decoder Approach in two stages. First, we performed segmentation and edge detection in a general way without pretraining the encoder with SSL and obtained the results (referred to as Approach 1). In the second approach, we transferred the pre-trained weights of the SimSiam model to the shared encoder and performed the segmentation and edge detection as downstream tasks to obtain the results (referred to as Approach 2). We have obtained 0.988 and 0.985 accuracy values for segmentation and edge detection for Approach 1 with an average Dice score value of 0.821. In the second approach, we came up with 0.981 and 0.979 accuracy values and an average Dice score of 0.859. For Approach 2, we tried using different training fractions of the BraTS 2018 dataset to analyse the effect of the amount of data used for the downstream task. We present the segmentation and edge detection results obtained using Approach 1, along with the basic hyperparameters applied. The average validation Dice score and IoU score, which are crucial evaluation metrics, show higher values, indicating a better overlap of tumour subregions. Table 2 shows the results of different case evaluations for a batch size of 1. Each row represents the LR, training loss (TL), training accuracy (TA), validation accuracy (VA), validation dice score (VDS), and validataion IoU score (VIoU) for the segmentation (Seg) and edge detection (ED) in Approach 1 and Approach 2.

Table 2. Results of different approaches

Case	LR	TL	TA	VA	VDS	VIoU
1-Seg	0.0001	0.2799	0.9920	0.9883	0.9879	0.7444
1-ED	0.0001	0.0155	0.9877	0.9849	–	–
2-Seg	0.0001	0.2816	0.9917	0.9811	0.9808	0.6401
2-ED	0.0001	0.0152	0.9885	0.9788	–	–

For Approach 2, we tried using different training fractions of the BraTS 2018 dataset to analyze the effect of the amount of data used for the downstream task and obtained the results in Table 3. By observing the dice score values, we can express that the model has performed considerably well even for 75% or 50% of the data, highlighting the impact of SSL. By increasing the training data from 50% to 75%, the model may have encountered a different distribution of data samples. The additional data could have introduced more variability and complexity that the model needed to learn, leading to a slight decrease in performance initially. We can see that when we increased the training data to 100%, the model's decisions have been generalised.

The visualization results of the segmentation branch obtained for both approaches 1 and 2 highlight the effectiveness of our architecture, where both approaches provide accurate segmentation masks compared to ground truth labels as shown in Figure 4. Among the two methods, Approach 2 has been able to catch more fine-grained tumour regions highlighting the importance of the combination of dual-decoder architecture and SSL. In addition to the segmentation masks, the edge masks generated by our models are nearly close to the original ground truth edge masks where both approaches provide almost identical boundary results of tumour subregions clearly as shown in Figure 5.

Table 3. Dice scores obtained for tumour sub-regions for different fractions

Data Fraction	Dice Score Values		
	Whole-Tumour	*Tumour-Core*	*Enhancing-Tumour*
100%	0.846	0.837	0.884
75%	0.705	0.723	0.668
50%	0.736	0.719	0.779

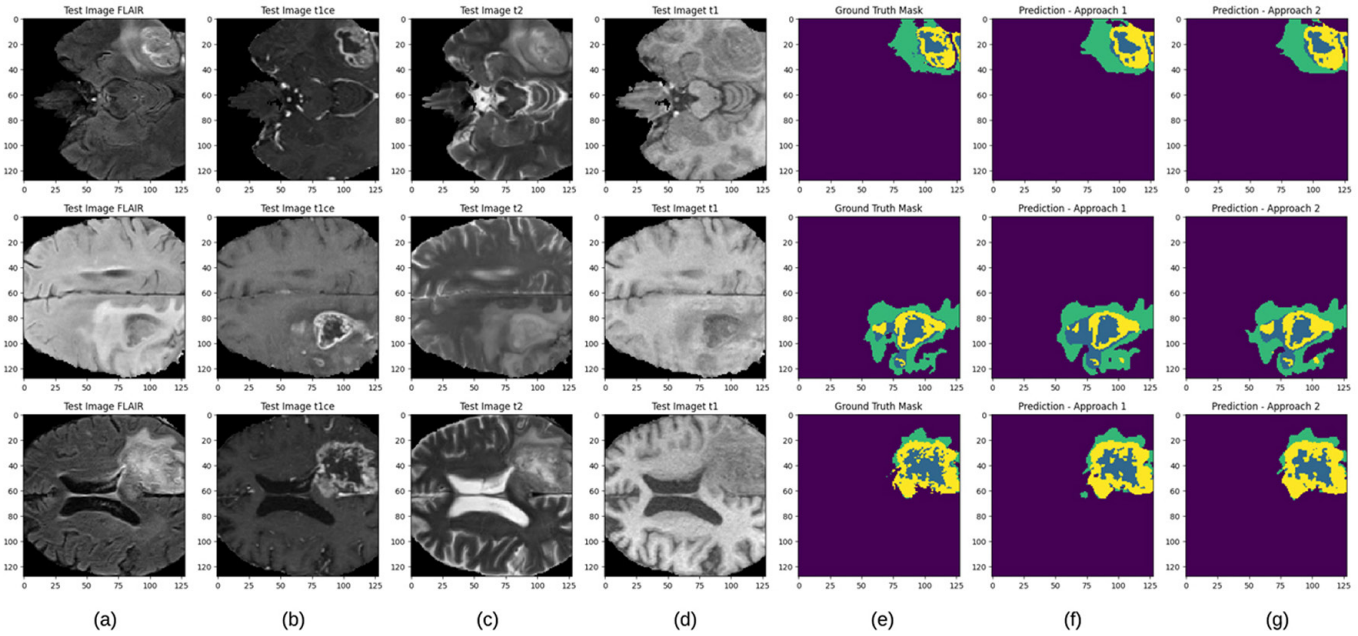

Fig. 4. Comparison of the visualization of the two approaches we followed – Figures from column (a) to (d) are the 3D MRI image inputs of 4 modalities, (e) represents original ground truth segmentation masks, (f) and (g) are the predictions of the Approach 1 and 2, respectively

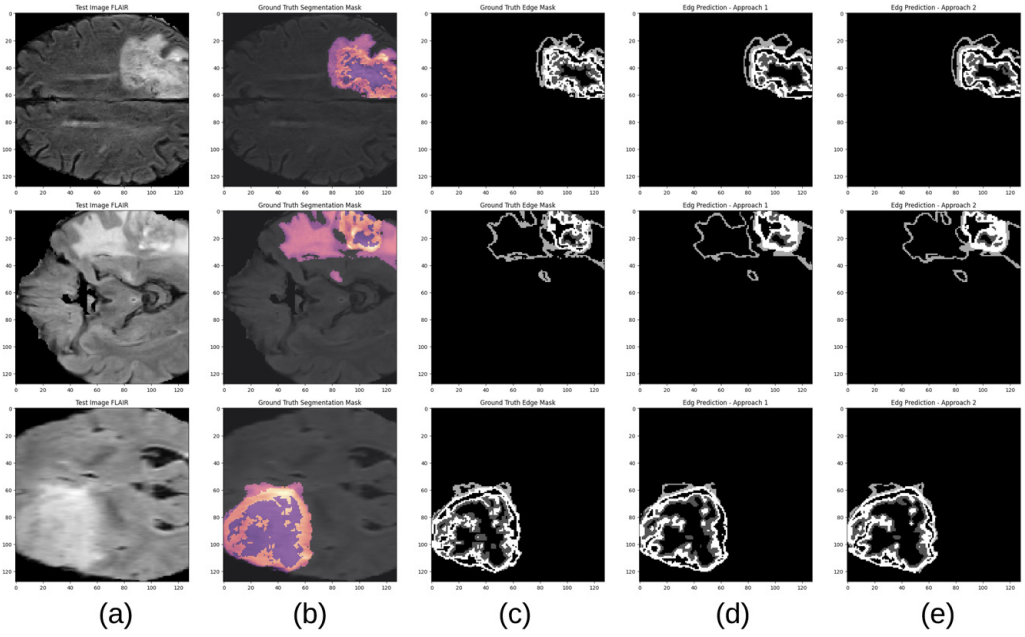

Fig. 5. Comparison of the edge detection process of the two approaches: Out of 4 input MRI modalities, column (a) shows the FLAIR input; column (b) is the segmentation output overlayed on the input; (c) represents the ground truth edge mask; (d) and (e) relate to edge outputs of Approaches 1 and 2, respectively

4.3 Comparison of different approaches

We performed a comprehensive comparison and analysis of our proposed model with the recent most effective brain tumour segmentation models. Since our approach consists of a self-supervised pretraining process and a novel dual-decoder architecture, we divided the comparison into two branches. As shown in Table 2, first, we compared our Approach 1 with major fully supervised model architectures, and the next comparison is our Approach 2 with recent self-supervised brain tumour segmentation approaches. Dice score values for each tumour subregion (WT, TC ET) have been used as the evaluation metric for all the comparisons as shown in Tables 4 and 5.

Table 4. Dice score value comparison of Approach 1 with fully supervised model

Model	BraTS Dataset	Dice Score Values		
		Whole-Tumour	*Tumour-Core*	*Enhancing-Tumour*
3D UNet [16]	2018	0.715	0.725	0.718
Dense-UNet [17]	2017	0.830	0.760	0.685
Inception-UNet [18]	2018	0.925	0.951	0.948
3D Attention-UNet [19]	2019	0.869	0.777	0.778
ResUNet [20]	2018	0.788	0.734	0.649
This study	**2018**	**0.840**	**0.811**	**0.811**

Table 5. Dice score value comparison of Approach 2 with self-supervised approaches

Model	BraTS Dataset	Dice Score Values		
		Whole-Tumour	*Tumour-Core*	*Enhancing-Tumour*
Layer Decomposition [21]	2018	0.901	0.824	0.773
Barlow Twins for UNet [9]	2018	0.651(Avg.)		
Barlow Twins for Att-UNet [9]	2018	0.769(Avg.)		
Hybrid Fusion [22]	2020	0.863	0.740	0.707
This study	**2020**	**0.846**	**0.837**	**0.884**

The results suggest that our novel dual-decoder segmentation approach has outperformed many supervised UNet variants except for more complex architectures like Inception UNet. This highlights the impact of boundary detection on the semantic segmentation of the tumours. In contrast to more complex UNet architectures, our model saves a considerable amount of computational demand, which is crucial in training 3D-CNN models. The results of our self-supervised approach are also satisfying, given the relatively simple and effective SimSiam architecture.

4.4 Existing challenges and future extensions

Our novel approach for brain tumour segmentation and edge detection shows indicative results. One of the biggest challenges of our research is the need for

labelled data for training and validation. While using multi-modal 3D images gives detailed information about brain structures, manually labelling such a large amount of MRI data is very time-consuming and not practical for clinicians. This challenge is made worse by differences in labelling accuracy among different clinicians, which can cause inconsistencies and biases in the dataset. Even though we try to use SSL models that work with unlabelled data, the initial training phase still needs a lot of labelled data to reach a basic level of accuracy.

Creating and training complex models, especially those using 3D image data, requires significant computational resources. High-performance GPUs, large memory capacities, and long processing times are needed to handle the intensive computations involved in 3D segmentation tasks. This need limits the accessibility of our approach, as not all research facilities or medical institutions have the necessary computational infrastructure. Also, the time needed to train such models can be very long, slowing down the development and use of practical diagnostic tools.

The proposed model has addressed challenges related to computational resources and real-time processing. With the learning strategies, the architecture supported to achieving high performance while maintaining lower resource demands, making it suitable for applications that require rapid processing and accurate segmentation. At its core, the implementation of the SimSiam model for SSL demonstrates a particularly efficient approach to memory management. By utilising only positive pairs for similarity learning rather than requiring negative pairs, our model significantly reduces memory requirements during training. This optimisation is especially crucial when working with 3D CNN models, where memory constraints can often become a limiting factor. The elimination of negative pairs not only conserves memory but also simplifies the training process for medical image segmentation, where generating appropriate negative pairs can be challenging. The architecture's design further reflects careful consideration of resource optimisation. Our model employs a shared encoder that serves both decoders (P and S), representing a measured choice to reduce the total number of parameters while maintaining model capability. This sharing of resources ensures efficient utilisation of computational capacity without compromising the model's ability to perform both edge detection and segmentation tasks. The integration of dropout for regularisation and max pooling for downsampling represents additional measures to maintain computational efficiency while preserving model performance. Additionally, SimSiam is an SSL framework that allows the model to learn representations without requiring extensive labelled data. The use of efficient pooling strategies helps reduce inference time. Further, the model design allows for parallel processing of multiple input frames, which enhances throughput during real-time applications.

While our approach aims to improve segmentation accuracy and explainability, making the model work well across different patient groups remains a challenge. Brain tumour characteristics can vary widely among patients, influenced by factors like age, genetic background, and tumour type. A model trained on a specific dataset may not perform equally well on new, unseen data from different populations. Ensuring the model is robust enough to handle such variability requires extensive testing and possibly additional training on more diverse datasets, which is not always possible.

Even though we include explainable AI (XAI) techniques to make the model's decision-making process more transparent, balancing high performance and explainability is challenging. Some of the most advanced deep learning models, like those used for image segmentation, are very complex and can be hard to interpret.

While edge detection and other XAI methods help provide insights into the model's outputs, there is still a gap in fully understanding and explaining how these models work to medical professionals, which is important for clinical use.

Applying our research findings to real-world applications presents several logistical and technical challenges. Healthcare systems need new technologies to fit seamlessly into existing workflows, which involves compatibility with current software, hardware, and data management systems. Ensuring the reliability and accuracy of AI-driven diagnostic tools in a clinical setting also requires rigorous validation and certification processes, which can be time-consuming and expensive. Additionally, training and gaining acceptance from medical professionals are essential for successful implementation, requiring comprehensive education and support programs.

Moreover, we have identified potential future extensions that can enhance the effectiveness of our research. Incorporating explainable AI (XAI) techniques would provide valuable insights into the model's decision-making process, establishing trust in clinicians [23], [24]. Since our approach does not have a tumour classification process, a typical explainability application will not create any useful impact on our model's interpretability. Therefore, moving towards a post-hoc XAI approach for segmentation would be more effective.

The current architecture utilizes separate decoders for segmentation and edge detection. Investigating connections between these decoders could potentially improve performance by facilitating information exchange between two decoder branches. Thus, the proposed approach can be future enhanced feature sharing or cross-attention mechanisms between decoders. This leverages the strengths of multiple modalities and attention-based techniques to improve segmentation accuracy. Since feature sharing exchanges the learned features between different decoders, each decoder can benefit from the information captured by others. This method can enhance the model's ability to generalize across various tumor characteristics and imaging modalities. Similarly, cross-attention allows one decoder to focus on relevant features from another decoder's output. This is useful in multi-modal imaging scenarios where different modalities provide unique insights about the tumor. Thus, the segmentation model dynamically prioritizes features that are most relevant for accurate segmentation, thus improving overall performance, even with less training data while reducing the computational cost

Moreover, implementing a user-friendly application as a support tool for clinical purposes would be a practical approach to utilise this study [25]. The development of such applications should strongly focus on integrating into hospital systems, analysing compatibility with available MRI data inputs, and ensuring the data privacy of patients.

5 CONCLUSION

This paper introduced a comprehensive approach to boundary-focused brain tumour segmentation with SSL. We applied a 3D SimSiam network architecture to implement a contrastive learning-based approach to pretrain an encoder. Our dual-decoder architecture is a novel approach for the downstream segmentation task that focuses on improving both the segmentation accuracy of tumour regions and edge detection accuracy of the tumour boundaries. Our model generates the segmentation mask and the edge mask separately, providing comprehensive and critical details for clinicians that are not always present in typical segmentation results. Compared to other recent SSL models, our 3D SimSiam network provides

a simpler model architecture that requires lower computation demand and training time. The exclusion of negative pairs in this contrastive learning approach is another significant improvement of our study, which is highly impactful in the medical image domain. The combination of brain tumour segmentation and edge detection is the most impactful contribution we have presented in this study. We have proven that the additional edge detection task of the dual-decoder 3D-Unet model has been able to improve the segmentation results, highlighting the impact of our novel model architecture. Moreover, the additional edge mask output has a significant clinical value. According to clinicians, usually, radiotherapy treatments are preferred over surgeries in brain tumour treatments due to the risks associated with brain surgeries. Accurate identification of tumuor boundaries is essential for radiation oncologists to precisely target the radiation beams to the tumour volume while sparing surrounding healthy brain tissue. This is where the edge mask output provides detailed information to support the radiation treatment process. We obtained maximum accuracy values of 0.988 and 0.985 for segmentation and edge detection, respectively, and maximum segmentation dice scores of 0.846, 0.837, and 0.884 for WT, TC, and ET regions, respectively.

6 REFERENCES

[1] N. Wijethilake, D. Meedeniya, C. Chitraranjan, I. Perera, M. Islam, and H. Ren, "Glioma survival analysis empowered with data engineering—a survey," *IEEE Access*, vol. 9, pp. 43168–43191, 2021. https://doi.org/10.1109/ACCESS.2021.3065965

[2] S. Dasanayaka, V. Shantha, S. Silva, T. Ambegoda, and D. Meedeniya, "Interpretable machine learning for brain tumour analysis using MRI," in *2022 2nd International Conference on Advanced Research in Computing (ICARC)*, 2022, pp. 212–217. https://doi.org/10.1109/ICARC54489.2022.9754131

[3] S. Dasanayaka, V. Shantha, S. Silva, D. Meedeniya, and T. Ambegoda, "Interpretable machine learning for brain tumour analysis using MRI and whole slide images," *Software Impacts*, vol. 13, p. 100340, 2022. https://doi.org/10.1016/j.simpa.2022.100340

[4] A. Taleb *et al.*, "3D self-supervised methods for medical imaging," *Advances in Neural Information Processing Systems*, vol. 33, pp. 18 158–18 172, 2020.

[5] J. Wu, J. Ma, H. Xi, J. Li, and J. Zhu, "Multi-scale graph harmonies: Unleashing U-Net's potential for medical image segmentation through contrastive learning," *Neural Networks*, vol. 182, p. 106914, 2025. https://doi.org/10.1016/j.neunet.2024.106914

[6] W. J. Zhang, W. T. Chen, C. H. Liu, S. W. Chen, Y. H. Lai, and S. D. You, "Feasibility study of detecting and segmenting small brain tumors in a small MRI dataset with self-supervised learning," *Diagnostics*, vol. 15, no. 3, p. 249, 2025. https://doi.org/10.3390/diagnostics15030249

[7] P. Yan *et al.*, "Online self-distillation and self-modeling for 3D brain tumor segmentation," *IEEE Journal of Biomedical and Health Informatics,* 2025.

[8] X. Xu, J. Chen, D. Thakur, and D. Hong, "Multi-modal disease segmentation with continual learning and adaptive decision fusion," *Information Fusion*, vol. 118, p. 102962, 2025. https://doi.org/10.1016/j.inffus.2025.102962

[9] N. S. Punn and S. Agarwal, "Bt-UNet: A self-supervised learning framework for biomedical image segmentation using barlow twins with U-net models," *Machine Learning*, vol. 111, pp. 4585–4600, 2022. https://doi.org/10.1007/s10994-022-06219-3

[10] P. K. Sathish, "Advancing brain tumor segmentation in MRI scans: Hybrid attention-residual UNET with transformer blocks," *International Journal of Online & Biomedical Engineering*, vol. 20, no. 6, pp. 103–115, 2024. https://doi.org/10.3991/ijoe.v20i06.46979

[11] G. L. E. Maquen-Niño *et al.*, "Brain tumor classification deep learning model using neural networks," *International Journal of Online & Biomedical Engineering*, vol. 19, no. 9, pp. 81–92, 2023. https://doi.org/10.3991/ijoe.v19i09.38819

[12] S. M. Mangj, P. H. Hussan, and W. M. R. Shakir, "Efficient deep learning approach for detection of brain tumor disease," *International Journal of Online & Biomedical Engineering*, vol. 19, no. 6, pp. 66–80, 2023. https://doi.org/10.3991/ijoe.v19i06.40277

[13] "Brats challenge." http://www.braintumoursegmentation.org/ [Accessed: Jun. 05, 2023].

[14] D. Meedeniya, *Deep Learning: A Beginners' Guide*. Chapman & Hall, 2023. [Online]. Available: www.routledge.com/9781032473246

[15] S. Sahayam and U. Jayaraman, "Integrating edges into U-net models with explainable activation maps for brain tumour segmentation using MR images," *arXiv preprint arXiv: 2401.01303*, 2024.

[16] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, 2016, pp. 424–432. https://doi.org/10.1007/978-3-319-46723-8_49

[17] M. Shaikh, G. Anand, G. Acharya, A. Amrutkar, V. Alex, and G. Krishnamurthi, "Brain tumour segmentation using dense fully convolutional neural network," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, 2018, pp. 309–319. https://doi.org/10.1007/978-3-319-75238-9_27

[18] A. Waqas, D. Dera, G. Rasool, N. C. Bouaynaya, and H. M. Fathallah-Shaykh, "Brain tumour segmentation and surveillance with deep artificial neural networks," *Deep Learning for Biomedical Data Analysis*, 2021, pp. 311–350. https://doi.org/10.1007/978-3-030-71676-9_13

[19] M. Islam, V. Vibashan, V. J. M. Jose, N. Wijethilake, U. Utkarsh, and H. Ren, "Brain tumour segmentation and survival prediction using 3D attention UNet," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, vol. 11992, 2020, pp. 262–272. https://doi.org/10.1007/978-3-030-46640-4_25

[20] A. Kermi, I. Mahmoudi, and M. T. Khadir, "Deep convolutional neural networks using U-Net for automatic brain tumour segmentation in multimodal MRI volumes," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, 2019, vol. 11384, pp. 37–48. https://doi.org/10.1007/978-3-030-11726-9_4

[21] X. Zhang *et al.*, "Self-supervised tumour segmentation through layer decomposition," *arXiv preprint arXiv:2109.03230*, 2021.

[22] L. Zhao, C. Jia, J. Ma, Y. Shao, Z. Liu, and H. Yuan, "Medical image segmentation based on self-supervised hybrid fusion network," *Frontiers in Oncology*, vol. 13, p. 1109786, 2023. https://doi.org/10.3389/fonc.2023.1109786

[23] L. Gamage, U. Isuranga, D. Meedeniya, S. De Silva, and P. Yogarajah, "Melanoma skin cancer identification with explainability utilizing mask guided technique," *Electronics*, vol. 13, no. 4, p. 680, 2024. https://doi.org/10.3390/electronics13040680

[24] N. Sritharan, N. Gnanavel, P. Inparaj, D. Meedeniya, and P. Yogarajah, "EnsembleCAM: Unified visualization for explainable cervical cancer identification," in *2024 International Research Conference on Smart Computing and Systems Engineering*, 2024, pp. 1–6. https://doi.org/10.1109/SCSE61872.2024.10550859

[25] T. Shyamalee, D. Meedeniya, G. Lim, and M. Karunarathne, "Automated tool support for glaucoma identification with explainability using fundus images," *IEEE Access*, vol. 12, pp. 17290–17307, 2024. https://doi.org/10.1109/ACCESS.2024.3359698

7 AUTHORS

Dasith Samarasinghe holds a BSc in Computer Science and Engineering, University of Moratuwa, Sri Lanka. His research interests span deep learning and software engineering.

Deshan Wickramasinghe holds a BSc in Computer Science and Engineering, University of Moratuwa, Sri Lanka. His research interests span deep learning and software engineering.

Theshan Wijerathne holds a BSc in Computer Science and Engineering, University of Moratuwa, Sri Lanka. His research interests span deep learning and software engineering.

Dulani Meedeniya is a Professor in Computer Science and Engineering at the University of Moratuwa, Sri Lanka. She holds a PhD in Computer Science from the University of St Andrews, United Kingdom. She is the Director of the Bio-Health Informatics group and Eco-Sustainable Informatics group at her department. She is a Fellow of HEA (UK), MIET, Senior Member IEEE, Member of ACM and a Chartered Engineer registered at EC (UK).

Pratheepan Yogarajah is a Lecturer in Computing Science at the Ulster University. He obtained his MPhil in Computer Vision from the Oxford Brookes University, UK, and PhD from the Ulster University, UK. His research interests include biometrics, computer vision, image processing, steganography and digital watermarking, and machine learning (E-mail: p.yogarajah@ulster.ac.uk).