International Journal of
Computer Science and Artificial Intelligence
ISSN: 3143-6986 | https://www.ioro.org/ijcsa

BRAIN TUMOR SEGMENTATION OF MRI SEQUENCES (T1, T2, T1CE, FLAIR) USING BRATS DATASET

Lovedeep Kaur [1], Er. Parminder Singh [2], Dr. Naveen Dhillon [3]

[1] *M.Tech Research scholar at Ramgarhia Institute of Engineering and Technology, Phagwara, Punjab, india*

[2] *Head of Computer Science Department at Ramgarhia Institute Of Engineering and Technology, Phagwara, Punjab*

[3] *Principal at Ramgarhia Institute Of Engineering and Technology, Phagwara, Punjab, india*

* **Corresponding author: Lovedeepkaur1991@gmail.com**

Abstract

Brain tumor segmentation from Magnetic Resonance Imaging (MRI) is an important task in computer-aided diagnosis because accurate identification of tumor regions supports clinical assessment and treatment planning. However, the complex structure, irregular shape, intensity variation, and heterogeneous appearance of brain tumors make automated segmentation challenging. This study presents a comparative deep learning framework for brain tumor segmentation using the BraTS 2020 dataset and two-dimensional (2D) MRI images. In this study, we explore 2D deep learning architectures for automated tumor segmentation, focusing on U-Net, Vision Transformer (ViT), and Res-ViT model. U-Net, with its encoder–decoder design and skip connections, has been widely adopted for medical image segmentation due to its ability to capture fine-grained spatial features. ViT, leveraging self-attention mechanisms, introduces a global receptive field that enhances contextual understanding across slices. The Res-ViT hybrid combines residual learning with transformer-based attention, aiming to balance local feature extraction and long-range dependency modeling. Preprocessing steps, including skull stripping, intensity normalization, and bias field correction, were applied to ensure consistency across scans. Data augmentation techniques such as rotation, flipping, and elastic deformation were employed to mitigate overfitting and improve generalization. The models are evaluated using important segmentation metrics, including, Intersection over Union (IoU), accuracy, precision, recall/sensitivity, loss, and 95th-percentile Hausdorff Distance (HD95). The comparative analysis aims to identify the strengths and limitations of convolutional and transformer-based approaches for 2D brain tumor segmentation. The study demonstrates the potential of combining local feature extraction and global contextual learning to achieve more accurate and robust brain tumor segmentation from multimodal MRI images.

Keywords: MRI, U-Net, ViT, BraTS 2020, IOU, HD95.

1. Introduction

Brain Tumors present one of the greatest challenges in contemporary medicine, due to their complexity, variation, and high mortality rates. They can be generally classified into two main types, i.e., Benign Tumors & Malignant Tumors. Malignant Tumors, such as glioblastomas, present the greatest risk for patient outcomes. Therefore, a detailed, accurate, and timely diagnosis of Brain tumor types is necessary to guide appropriate approaches to treatment for patients, including interventions through Surgery, Radiation, and/or Chemotherapy [1]. The current "gold standard" for Brain Tumor Diagnosis is MRI, due to its non-invasive and very high-resolution images of Soft Tissues in the Brain. However, an interpretation of an MRI requires specialized knowledge, and the process for manual delineation of the tumor volumes is labor-intensive, and

International Journal of Computer Science and Artificial Intelligence (IJCSA)
Volume 1 | Issue 1 | 2026 | Article ID: 2800-1794-9912 | https://doi.org/10.64823/ijcsa.2601010
IORO Publications · Texas, USA · www.ioro.org

there is an inherent inter-observer variability among radiologists when it comes to delineating tumor volumes [2].

Medical Image Analysis is being revolutionized by the development of Artificial Intelligence (AI) and more specifically, Deep Learning (DL). This also applies to the segmentation of brain tumors, where these models can provide high-quality (accurate) tumor region delineation to radiologists in an automated manner [3]. However, the quality of DL supervised models depends on the availability of large amounts of annotated data. Unfortunately, there is a shortage of annotated datasets within the field of medical imaging due to the highly specialist nature of labelling and the difficulty and costs of the annotation process itself [4].

Due to this problem, Researchers have begun to experiment with Self-Supervised Learning (SSL) as an up-and-coming solution [5]. Through SSL, the model learns to extract features from a multitude of unlabeled images (which can easily be obtained) and, following model training, learns using fewer labelled images [6]. SSL has thus reduced the need to generate labelled images, one of the major impediments to the development of medical AI. The motivation for doing this research is to tap into the vast underutilized stores of unlabeled Cnn MRI datasets [7].

Moreover, by combining multiple MRI modalities (i.e., T1, T2, T1c, and Fluid-Attenuated Inversion Recovery (FLAIR)), the reliability of brain tumor segmentation is improved because each modality offers additional value through the merging of clinical data from each modality [8]. As interest increases in multi-modal approaches to learning, so too does the limited implementation of semi-supervised learning (SSL) paradigms into multi-modal frameworks for the segmentation of brain tumors [9].

Brain Tumor Segmentation has become quite prevalent within contemporary medical diagnostics and treatment planning, as evidenced by [10]. Accurate identification of tumor regions during diagnostic MRI scans is critical for determining not only if surgical resection is an option, but also where radiation therapy should be aimed and what chemotherapy choice would be best for the patient. Resultantly, accurate and reliable segmentation of the brain tumor must be of utmost importance in ensuring a safe and effective clinical operation for the patient [11].

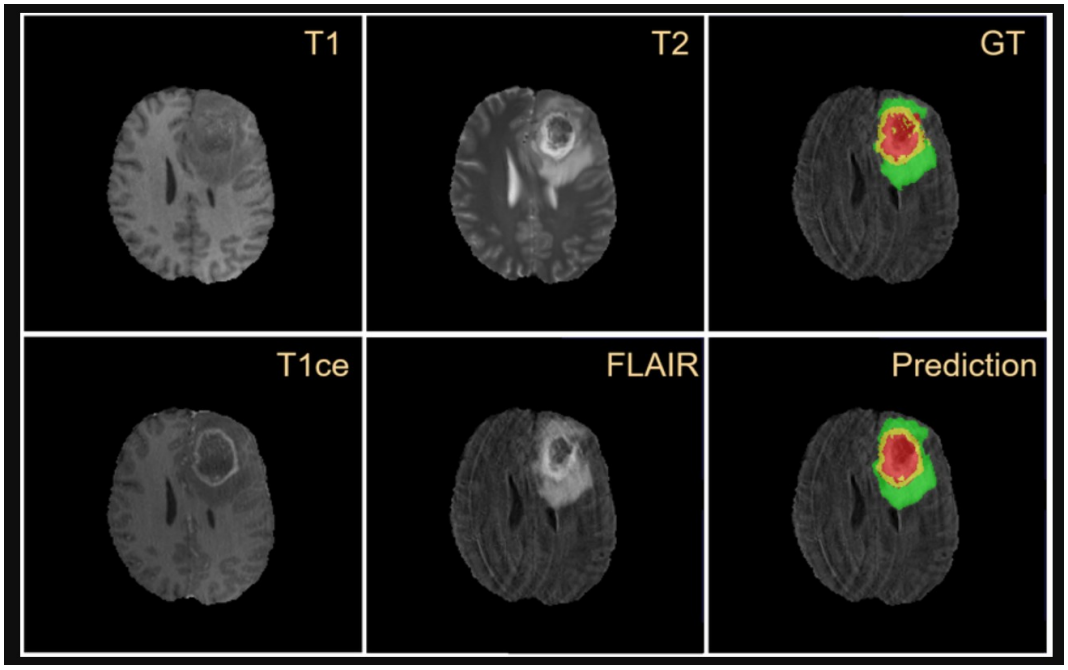

Figure 1: Multiple MRI modalities—such as T1, T2, T1ce, and FLAIR

MRI imaging is relevant to the second aspect of the significance of MRI, i.e., the multiple modalities involved with MRI imaging. Each MRI sequence does not reflect the complete complexity of a brain tumor [12]. For example, T1-weighted scans provide structural detail, while T2-weighted scans depict areas of edema. In combining the aforementioned multimodality, the combined use of T1, T2, T1c, and FLAIR will allow for the production of more accurate representations of tumor histology and will aid in deriving segmentation that reflects both anatomical and pathological factors [13].

Different types of MRI sequences show different aspects of the different types of anatomy and pathology associated with MRI; each therefore plays a complementary role when combined into a single analysis.

- The T1-weighted sequences are most useful for evaluating the structure of the brain's anatomy, as T1-weighted scans give detailed structural information about the brain, providing the means to visually distinguish between Gray and White Matter within the brain.
- The T2-weighted sequences: Highlight areas of high-water content, making T2-weighted scans most useful for identifying edema or fluid associated with tumor cystic components.
- T1c (Contrast-enhanced T1): Following the administration of an intravenous contrast agent, regions of active tumor, which absorb the agent, become more apparent and visually stand out from other normal surrounding structures by showing a clear definition of tumor margins.
- FLAIR: Provides a suppression of fluid signals, making it easier to visualize any lesions present in a region of cerebrospinal fluid.

The combination of all these modalities allows you to see a holistic view when you use all these modalities together. An example of this is that T2 can identify edema, whereas FLAIR can show whether or not there is a tumor located in the area, and not only cerebrospinal fluid [14]. Additionally, T1c, which has been enhanced with contrast, can help differentiate between active tumor tissue and necrotic/non-enhancing tissue.

The segmentation accuracy is also necessary for the quantitative measurement of tumor growth (i.e., by having the ability to compare tumor volumes from different points in time, clinicians will be able to observe how effective their treatments are and potentially identify a recurrence sooner) [15]. Subsequently, automated techniques that create consistent and high-quality tumor edges can aid longitudinal research, clinical trials, and individualized treatment planning. The segmentation of brain tumors can also be applied beyond the field of brain tumor research and treatment. Automated tumor segmentation reduces the burden on clinicians for annotating data and enables the generation of large, annotated datasets with minimal manual effort [16].

Automated analysis of medical images has made progress but brain tumor segmentation still faces a number of challenges, particularly in low-labeled MRI situations. This can be time-consuming, expensive, and often impractical at scale. Because of this, most available brain tumor datasets are much smaller than datasets from the natural image domain and therefore supervised deep learning models will generalize poorly [17].

Variability in tumors presents a substantial problem, as there is a high variance in the way that brain tumors look. They come in all different sizes, shapes, locations, and intensity profiles. Infiltration of healthy tissue into some tumors makes it extremely difficult to determine where the tumor ends and the healthy tissue starts, even for highly trained experts [18]. Tumors found in pediatric patients differ markedly from those found in adults and introduce additional challenges when trying to generalize from one set of data to another. Because of this variability, it is common for models trained on a particular dataset to perform poorly when presented

with clinical data from a different source (in situations where there are differences in imaging protocols or scanner settings) [19].

There is further complexity brought about by the multimodality of MRI. Although T1, T2, T1c, and FLAIR help provide more comprehensive characterization of the tumor, the combination of the modalities would need specific attention in preprocessing, alignment, and normalization. One of the biggest challenges is to develop powerful models that can deal with incomplete or noisy multimodal data [20].

To conclude, low-label MRI environments pose several problems of complexity: annotated data is limited, tumors are heterogeneous, multi-modal integration problems arise, classes are imbalanced, and implementation considerations are challenged [21]. The conventional supervised deep learning is highly dependent on big annotated databases. This can be a significant limitation in medical imaging, as annotations of high quality may take a great amount of manual work and require radiological experience. This gap has stimulated the research of Self-Supervised Learning (SSL) as a potential paradigm of brain tumor segmentation [22].

SSL uses unlabeled data to pretrain models using auxiliary tasks or otherwise referred to as pretext tasks. These are not those tasks that need manual annotations but can also enable models to learn useful feature representations. It can be used in predicting missing pieces of an image, solving jigsaw-like puzzles of image fragments, or comparing various augmentations of the same scan [23]. The first benefit of the use of the SSL in medical imaging is that there is a high number of unlabeled MRI scans. Modelling Learners of universal imaging features and their transfer across datasets and modalities, through pretraining on large unlabeled datasets. This decreases overfitting and increases generalization to missing clinical situations [24].

The inclusion of the SSL in the multi-modal segmentation models enhances even more. Through the combination learning of various MRI sequences, the SSL can learn the cross-modality connection that enhances the accuracy of tumor delineation. This becomes particularly crucial when the appearance of tumors in different modalities is different.

2. Literature Review

Multimodal MRI segmentation uses different MRI types (T1, T2, T1ce, FLAIR) together to get richer information for tasks like brain tumors. We will apply deep learning and advanced fusion techniques for better accuracy, handling noise. The work has been done to discover the issues associated with MRI of brain tumors. This literature survey attempts to analyses further methods and explore the opportunities to develop the best possible solution using new technologies. Some include the following:

Saha et al., 2024. The Magnetic Resonance Imaging (MRI) has stood out as the most important in the diagnosis of brain tumors since it has the best capacity of giving the soft tissues contrast and multi-parameters with no ionizing radiation. Every MRI modality has its own facets of anatomical and pathological detail of the brain. T1-weighted (T1) scans are needed to underline anatomical features and identify the Gray and white matter, whereas T2-weighted (T2) scans are necessary to highlight the presence of fluid, which makes edema and tumor margins more visible. Contrast agents are applied during T1 using contrast (T1c), which allows better visualization of areas of blood-brain barrier disruption, which is characteristic of high-grade gliomas. FLAIR (Fluid-Attenuated Inversion Recovery) sequences inhibit cerebrospinal fluid, enhancing the definition of peritumoral edema [25].

Huang el al 2025 this paper talks about how multimodal foundation models are being used in medical imaging and how they can be applied in real healthcare settings. Multimodal foundation models are advanced AI

models that can learn from different types of data at the same time, such as medical images, doctor's report notes, clinical reports and patient records. Instead of relying on only one type of data, these models combine multiple sources to give better and more accurate results for prediction [26].

Pham et al., 2000 this paper screams the 1144 studies published between 2012 and 2024 and closely analyzed 48 key papers Early works utilized intensity-based thresholding and region growing, where tumor regions were identified based on pixel intensity distributions. Later, clustering techniques such as k-means and fuzzy c-means (FCM) were applied. An Adaptive Fuzzy C-Means (A-FCM), using spatial information, was proposed to increase robustness towards noise but not for irregular boundary tumors [27].

Seghier et al, 2008 it supervised machine learning approaches, using hand-crafted features, also emerged. Texture-based features were used to perform Support Vector Machines (SVM), Random Forest, and k-Nearest Neighbour (k-NN) Classification Models. These models showed promise; however, due to their reliance on manual feature engineering, they are difficult to scale and generalize and cannot represent the heterogeneity of tumors [28].

Tsangaratos et al, 2025 Therefore, Support Vector Machines (SVM), Random Forest, and k-Nearest Neighbour (k-NN) techniques are now used as a means of producing multi-level hierarchical feature representations that require no manual input (). Traditional methods continue to be useful in the preprocessing or as part of a hybrid system for a semi-automatic segmentation workflow, while deep learning-based techniques offer the best overall performance at this time [29].

Sarvamangala et al. (2022), Deep learning has brought significant improvements to the field of medical imaging and image segmentation, making it possible to replace traditional methods; for example, Convolutional Neural Networks (CNN's), have become the popular choice to solve this problem, as CNN's do not require the manual creation of features, Variants of U-Net have been the primary architecture for performing medical segmentation tasks, such as Brain Tumor Analysis [30].

Havaei et al. (2023) For instance, in the Brain Tumor Segmentation task, employed a Cascaded CNN approach to perform pixel-wise Classification, and their results were significantly better than those from other routine methods. In this paper presents a benchmark to evaluate domain adaption methods for brain MRI segmentation using T1-weighted images. MRI scans collected from different hospitals and scanners often vary in quality and appearance, which reduce the accuracy of segmentation models is called domain. To solve this problem the use of brain MRI data collected multiple imaging site [31].

Zhang, 2023 Deep learning approaches are proven to be more accurate than conventional treatment methods and exhibit greater resilience to the effects of Tumor heterogeneity. However, deep learning models require large amounts of labelled training data, which creates a significant barrier to developing deep learning models for medical imaging. This has led researchers to explore strategies to enhance deep learning's ability to learn from limited training datasets by utilizing data augmentation, transfer learning, and semi-supervised learning [32].

Bonato et al., 2025 Though deep learning employs very well-designed algorithms, researchers still contend that these approaches have limited interpretability, suffer from domain shifts across different scanners, and require extensive computational resources. Nonetheless, deep learning has had an important place in brain Tumor segmentation studies, allowing for the continued development of multimodal learning and self-supervised learning technique[33].

Valverde et al. (2021) multi-modal deep learning frameworks are intended to produce a single model utilizing multiple magnetic resonance imaging (MRI) modalities (T1, T2, T1c, FLAIR) to take advantage of the unique

features found in each modality. For the first time leveraged the BraTS dataset to emphasize the importance of a multi-modal approach, observing that Tumor subregions appeared differently across MRI modalities [34].

Havaei et al. (2017) A variety of architectures exist that support the integration of multiple MRI modalities. An early example of an approach that used early fusion (concatenation) of all modalities in the input layer was proposed by. This technique is relatively straightforward; however, it may not adequately take full advantage of the strengths of the individual modalities .it also proposed later fusion models, which fused information from different modality-specific branches at subsequent levels of processing, allowing for better representation of the specific characteristics of tumorous regions. This concept is further supported by Wang et al. (2025), who utilized multi-scale and multimodal convolutional neural networks in their DeepMedic project, establishing robustness through additional multi-modal CNNs [35].

3. Methodology

3.1. Research Gaps: -

Segmenting brain Tumors can be difficult due to the complex shapes of Tumors and the importance of accurate delineation. While current methods (traditional and deep learning) have improved how Tumors are segmented, there are still many limitations with the current approaches. This section identifies where these gaps exist.

- Computational Efficiency of Pre-training: Initial pre-training can be computationally expensive, which contradicts the requirements of low-level, resource-constrained environments that lack high-end GPUs. The development of more lightweight and computationally efficient SSL pretext tasks or
- Having a robust system that can support missing modalities means users can be confident that they are taking full advantage of using different MRI modalities to access a complete picture of what is happening with your patient. This is especially important when considering the varied types of information available from MRI modalities. While many current systems rely on simple concatenation and sharing of parameters, a more sophisticated approach is needed in order to adequately capture the disparate and complementary types of information from each modality.
- The variability in size, shape, and position for individual Tumors in the training set used for BraTS has resulted in a true challenge for those working on segmenting individual small and heterogeneous sub-regions of Tumors. Each sub-region has its own specific type of tissue, architecture, and extent; thus, accurately segmenting these entities, especially in cases where there are surgical artifacts and scans, remains difficult despite the widespread availability of these data types.
- Handling Class Imbalance and Fine Structures: Brain Tumors are complex, heterogeneous, and highly variable in shape, size, and location. Current models often struggle with class imbalance, where the smaller, clinically significant Tumor sub-regions (enhancing Tumor core, necrotic core) are often overlooked compared to the larger edema or whole brain. But New SSL methods are needed that explicitly emphasize edge-aware features and fine structural details, potentially using novel loss functions or training strategies to better handle the imbalanced and complex Tumor sub-compartments in the BraTS dataset.
- Optimizing the Fine-Tuning Phase: While SSL reduces the dependency on large annotated datasets, maximizing performance with only a small number of labelled samples during the fine-tuning phase remains a challenge. Research on optimal fine-tuning strategies that leverage the rich pre-learned

features from SSL with minimal labelled data to achieve performance comparable to fully supervised models is an active area for exploration.

3.2. Problem Formulation

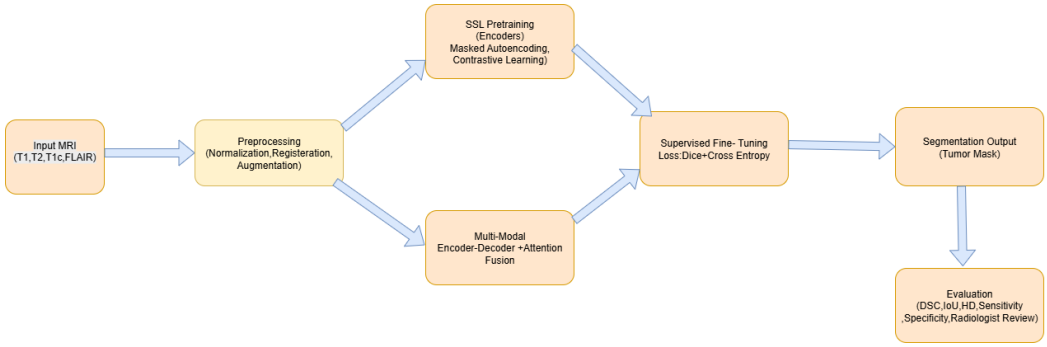

Figure2: Proposed methodology flowchart showing the complete pipeline — from multi-modal MRI input and preprocessing to self-supervised pretraining, attention-based encoder–decoder fusion, supervised fine-tuning, segmentation output, and final evaluation.

3.3. Research Methodology

The initial step in constructing any AI model is to collect, preprocess, and annotate the appropriate dataset(s) for training, testing, and validation. Generally, raw data is not usable as input for deep learning algorithms; therefore, you should only use structured, formatted datasets with annotations and high-quality images for deep learning. High-quality images include deleting poor-quality images, removing duplicate images, and utilizing augmentation techniques such as flipping, rotating, resizing, enlarging, and labelling augmented images.

After completing the previous procedure, we will construct a U-Net-like encoder-decoder architecture based on modality-specific encoders and a modality-based attention mechanism. The next step is to develop your model further. Each MRI modality, T1, T2, T1c, and FLAIR, is separately processed through a different encoder stream for the extraction of modality-specific features. These modality-specific features are then combined and weighted dynamically through the modality-attention mechanism to perform tumor segmentation. The decoder reconstructs the combined, weighted features to provide the final segmentation mask using skip connection layers to maintain spatial information. To limit overfitting, dropout and batch normalization layers are included within the model architecture.

The growth of automated MRI segmentation has provided new ways of utilizing large amounts of unlabeled information from Medical Image Databases. Rather than collecting Obstacles to Human Cost and Time associated with manual Annotation through the Radiologist route, it provides a pretexting mechanism for Learning to generate Feature representations, ultimately for use in downstream Processes, e.g., Segmenting Tumors from Brain MRIs. Since there is typically limited access to annotated neuro-imaging Data for Clinical Research, whereas access to unlabeled MRI images is very commonplace, Pretrained Models on unlabeled MRI can help easily.

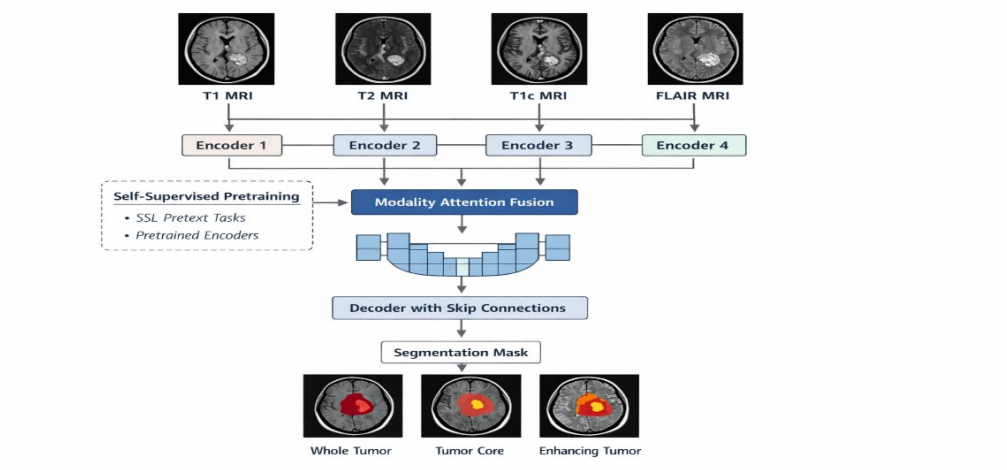

Figure 3: MRI segmentation with U-net Encoder and Decoder.

The Research focus is on how Transfer Learning can play an Important Role in building on the Successful Applications of SSL pre-training to provide a mechanism to allow the transferability of Learned Information from Pre-Trained SSL Models to Fine-Tune the Segmentation of Brain Tumors. Rather than building a New Model from Scarce Sources of Data and/or Labelled Data, a Fine-Tuning Approach to utilize Pre-Trained Encoder(s) can provide Rapid Convergence and Transfer Learning Capabilities to Allow Application Across Different Profiles (different Clinical Datasets) when These Profiles are considerably different from benchmark/Reference Dataset Sources, such as BraTS.

Evaluation Metrics:

1. Accuracy
2. Intersection over Union (IoU)
3. Hausdorff Distance (HD)
4. Precision
5. Recall
6. Loss

4. Results

4.1. Evaluation Metrics

4.1.1. Accuracy: This is just simply defined as how many of all the predictions were a correct prediction. While it is a useful figure, the number can create a foggy picture if a class is highly skewed.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

4.1.2. Precision: Precision is a measure of what percentage of those who test positive are actually positive, thus avoiding unnecessary worry or additional testing for healthy individuals.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

4.1.3. Recall (Sensitivity): High Recall - catching nearly all sick individuals, very few will be hidden.

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (3)$$

4.1.4. Specificity: High specificity ensures that most healthy individuals remain healthy and valuable limited resources are concentrated where care is needed.

$$Specificity = \frac{True\ Negative}{True\ Negative + False\ Positive} \quad (4)$$

4.1.5 IOU: - Intersection over Union (IoU) measures the degree of overlap between the predicted tumor region and the ground-truth tumor region.

IOU=

ALGORITHMS	Accuracy	Precision	Recall	IOU
U-NET	99.49%	90.53%	90.84%	75.64%
VIT	99.45%	89.00%	**93.82%**	**84.07%**
RES- VIT	99.45%	89.37%	93.39%	84.06%
HYBRID (U-NET)	**99.51%**	**90.55%**	91.51%	76.97%

$$\frac{True\ Positive}{True\ Positive + False\ Negative + False\ Postive} \quad (5)$$

Table1: Result of U-net, VIT, Res-net and Hybrid U-net

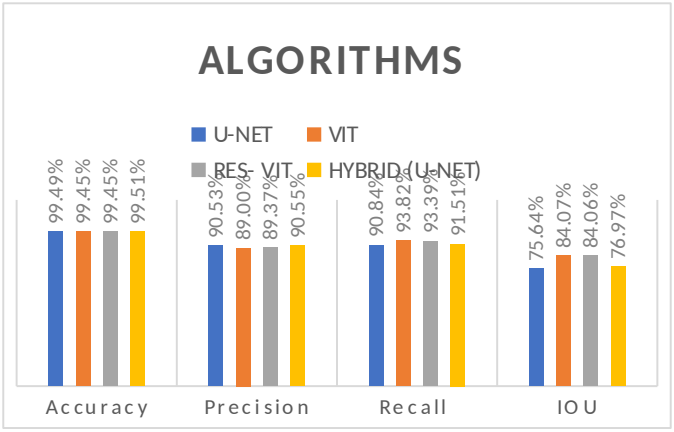

Figure 4: - Comparative performance of U-Net, ViT, Res-ViT, and Hybrid U-Net models in terms of accuracy, precision, recall, and IoU.

The table compares the performance of four MRI brain-tumor segmentation algorithms—U-Net, ViT, Res-ViT, and Hybrid U-Net—using Accuracy, Precision, Recall, and IoU. Overall, all four models achieve very high accuracy, with values above 99%. The Hybrid U-Net achieves the highest accuracy of 99.51%, slightly outperforming U-Net (99.49%), ViT (99.45%), and Res-ViT (99.45%). This indicates that the hybrid architecture provides slightly better overall pixel-level classification.

For Precision, the Hybrid U-Net obtains the highest value (90.55%), followed closely by U-Net (90.53%). Res-ViT achieves 89.37%, while ViT has the lowest precision at 89.00%. Higher precision indicates that the model produces fewer false-positive tumor predictions.

In terms of Recall, ViT performs best with 93.82%, followed by Res-ViT (93.39%), Hybrid U-Net (91.51%), and U-Net (90.84%). Therefore, ViT is better at identifying actual tumor pixels and produces fewer false negatives.

The IoU (Intersection over Union) is particularly important for segmentation because it measures the overlap between the predicted tumor region and the ground-truth region. ViT achieves the highest IoU of 84.07%, followed very closely by Res-ViT (84.06%). Hybrid U-Net achieves 76.97%, while standard U-Net obtains 75.64%. Thus, although Hybrid U-Net gives the best accuracy and precision, ViT provides the strongest segmentation overlap according to IoU.

Overall, the results show that Hybrid U-Net improves accuracy and precision, whereas ViT and Res-ViT provide better recall and IoU. These results suggest that transformer-based architectures can capture broader contextual information and improve tumor-region segmentation compared with conventional U-Net.

ALGORITHM	LOSS
U-NET	0.0134%
VIT	0.0154%
RES- VIT	0.0153%
HYBRID (U-NET)	**0.0127%**

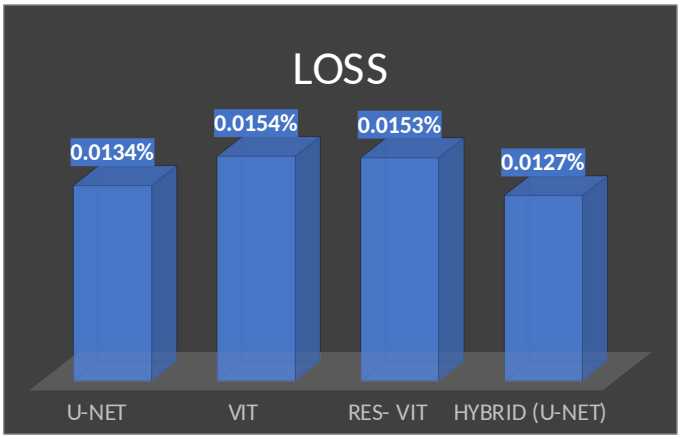

Table2 and Figure 5: Comparative analysis of loss values for different MRI brain tumor segmentation models.

The table compares the loss values of four MRI brain-tumor segmentation algorithms: U-Net, ViT, Res-ViT, and Hybrid U-Net. Loss represents the difference between the predicted segmentation and the ground-truth segmentation; therefore, a lower loss generally indicates better performance. The Hybrid U-Net achieves the lowest loss of 0.0127%, demonstrating the best performance among the compared models according to this metric. U-Net follows with a loss of 0.0134%, while Res-ViT and ViT obtain 0.0153% and 0.0154%, respectively. These results indicate that integrating advanced components into U-Net helps reduce prediction errors and improves segmentation performance. Thus, Hybrid U-Net provides the most favorable loss value.

ALGORITHM	HD95
U-NET	4.3985
VIT	3.8446
RES- VIT	3.1109
HYBRID (U-NET)	**2.7341**

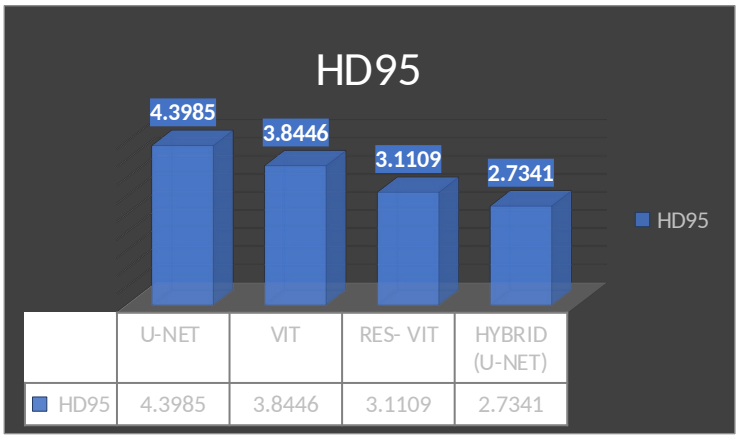

Table 3 and Figure 6: - Comparative analysis of HD95 values for U-Net, ViT, Res-ViT, and Hybrid U-Net models in MRI brain tumor segmentation.

The table compares the HD95 (95th percentile Hausdorff Distance) values of four MRI brain-tumor segmentation algorithms. HD95 measures the distance between the predicted tumor boundary and the ground-truth boundary. A lower HD95 value indicates better segmentation performance, as it means the

predicted boundaries are closer to the actual tumor boundaries. U-Net records an HD95 of 4.3985, while ViT achieves 3.8446. Res-ViT improves the result to 3.1109. The Hybrid U-Net achieves the lowest HD95 of 2.7341, making it the best-performing model for boundary accuracy. This improvement indicates that combining U-Net with advanced components provides more accurate and precise tumor boundary segmentation.

4.2. Visualization Techniques

Visuals were necessary to get the story going in the initial exploration.

4.2.1. Confusion Matrix: It summarised the predicted results for all categories and provided an insight about the prediction behaviour of the model, and unreliable and weak decision areas of the model.

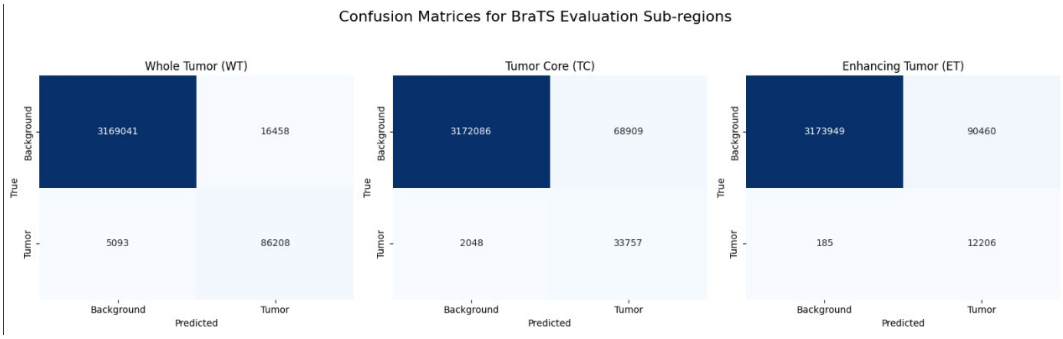

Figure 7: - Confusion matrix analysis of brain tumor segmentation performance across the BraTS sub-regions: Whole Tumor (WT), Tumor Core (TC), and Enhancing Tumor (ET).

4.2.2. ROC Curves: ROC curves for each contender allowed a visual checking of the comparison in sensitivity and specificity, and the area under each curve provided an easy scoreboard.

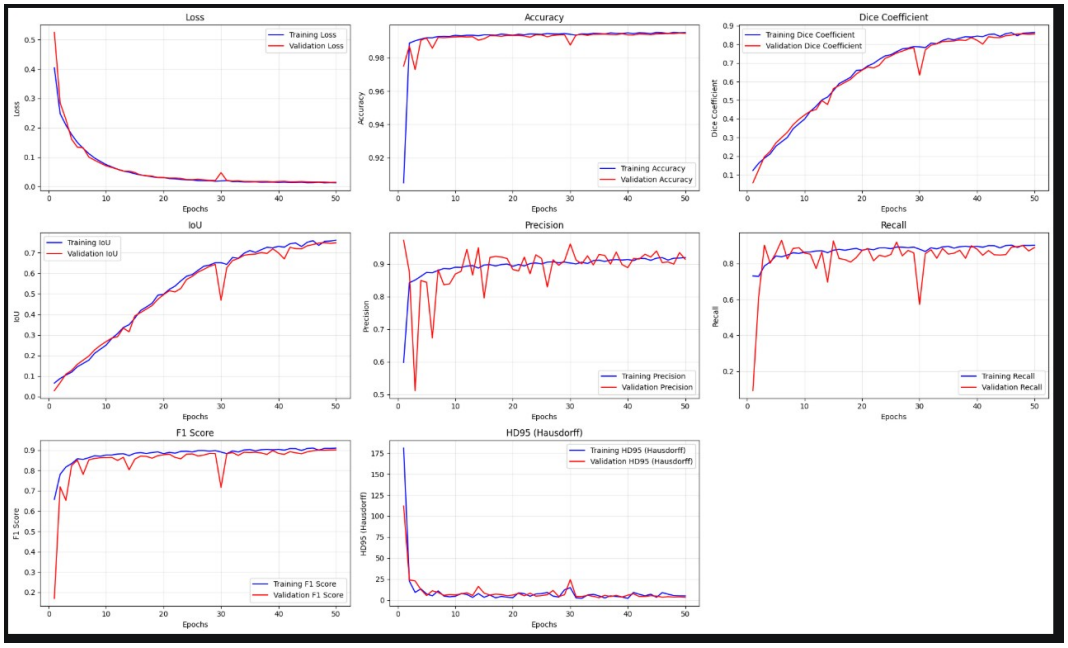

Figure 8: - ROC Curves performance of models in terms of accuracy, precision, recall, IOU, loss and HD95.

5. Conclusion

This study presented a comparative evaluation of four deep learning architectures—U-Net, Vision Transformer (ViT), Res-ViT, and Hybrid U-Net—for automated brain tumor segmentation on MRI data,

assessed across Accuracy, Precision, Recall, IoU, Loss, and HD95 boundary distance. The results collectively demonstrate that no single architecture dominates across every metric, and that the choice of "best" model depends heavily on which clinical or computational priority is being optimized.

In terms of overall pixel-level classification, all four models performed strongly, with accuracy exceeding 99% in every case. The Hybrid U-Net achieved the highest accuracy (99.51%) and the highest precision (90.55%), narrowly outperforming the standard U-Net.

However, when the focus shifts to sensitivity and spatial overlap—arguably the metrics most relevant to real-world diagnostic utility—the transformer-based architectures took the lead. ViT achieved the highest Recall (93.82%) and the highest IoU (84.07%), with Res-ViT close behind on both metrics. This indicates that self-attention-based architectures are better equipped to capture the broader contextual and long-range spatial relationships within MRI slices, allowing them to detect more true tumor pixels and produce segmentation masks that overlap more closely with the ground truth.

The loss and HD95 results add further nuance to this picture. The Hybrid U-Net recorded the lowest loss (0.0127%) and, notably, the lowest HD95 (2.7341), indicating that it produces the most accurate tumor boundary contours among the four models, even though its raw IoU score was lower than that of ViT and Res-ViT.

Taken together, these findings support the broader conclusion that hybrid architectures—which merge convolutional inductive biases with additional learned components—offer the best trade-off for metrics tied to pixel-level correctness (accuracy, precision) and boundary precision (loss, HD95), while pure and residual transformer architectures excel at holistic region detection and contextual understanding (recall, IoU). This reinforces the growing consensus in medical image segmentation literature that convolutional networks are strong local feature extractors, while transformer-based models, through self-attention, are better at modeling global spatial dependencies across the image. Neither approach is unconditionally superior; rather, their complementary strengths point toward the value of architectures that can jointly exploit both properties.

Future Scope

Several directions can extend and strengthen this work:

True hybrid transformer-convolutional fusion: Since the Hybrid U-Net already demonstrates the benefit of architectural fusion, a natural next step is to explicitly integrate transformer-based attention modules (as used in ViT and Res-ViT) directly into the U-Net encoder-decoder pathway—rather than as a separate comparison—to combine local feature precision with global contextual awareness within a single unified model. Architectures such as TransUNet, Swin-UNet, or attention-gated hybrid networks could be explored as logical extensions. Multi-modal and multi-sequence input fusion: Future work should investigate the impact of jointly leveraging all available MRI modalities (T1, T1ce, T2, FLAIR) through more sophisticated fusion strategies (e.g., cross-modal attention) rather than simple channel concatenation, potentially improving both recall and boundary precision simultaneously.

In summary, this research establishes a solid empirical foundation showing that architectural choice meaningfully shapes which aspect of segmentation performance is optimized. Future work aimed at architectural fusion, richer data utilization, boundary-aware training objectives, and clinical-grade validation will be critical to translating these promising results into deployable, trustworthy tools for brain tumor diagnosis and treatment planning.

Data Availability Statement

The BraTS 2020 Kaggle dataset offers multimodal, anonymized MRI scans with expert tumor annotations, enabling standardized brain tumor segmentation research.

References

[1] K. A. McNeill, "Epidemiology of brain tumors," *Neurol. Clin.*, vol. 34, no. 4, pp. 981–998, 2016.

[2] M. I. Razzak, M. Imran, and G. Xu, "Efficient brain tumor segmentation with multiscale two-pathway-group conventional neural networks," *IEEE J. Biomed. Heal. informatics*, vol. 23, no. 5, pp. 1911–1919, 2018.

[3] N. Arunkumar, M. A. Mohammed, S. A. Mostafa, D. A. Ibrahim, J. J. P. C. Rodrigues, and V. H. C. De Albuquerque, "Fully automatic model-based segmentation and classification approach for MRI brain tumor using artificial neural networks," *Concurr. Comput. Pract. Exp.*, vol. 32, no. 1, p. e4962, 2020.

[4] M. Aljabri, M. AlAmir, M. AlGhamdi, M. Abdel-Mottaleb, and F. Collado-Mesa, "Towards a better understanding of annotation tools for medical imaging: a survey," *Multimed. Tools Appl.*, vol. 81, no. 18, pp. 25877–25911, 2022.

[5] P. Kumar, P. Rawat, and S. Chauhan, "Contrastive self-supervised learning: review, progress, challenges and future research directions," *Int. J. Multimed. Inf. Retr.*, vol. 11, no. 4, pp. 461–488, 2022.

[6] Y. He *et al.*, "Foundation model for advancing healthcare: Challenges, opportunities and future directions," *IEEE Rev. Biomed. Eng.*, 2024.

[7] M. B. Hossain, R. K. Shinde, S. Oh, K.-C. Kwon, and N. Kim, "A systematic review and identification of the challenges of deep learning techniques for undersampled magnetic resonance image reconstruction," *Sensors*, vol. 24, no. 3, p. 753, 2024.

[8] R. Azad, N. Khosravi, M. Dehghanmanshadi, J. Cohen-Adad, and D. Merhof, "Medical image segmentation on mri images with missing modalities: A review," *arXiv Prepr. arXiv2203.06217*, 2022.

[9] C. Ambroise, "Structure-aware neural networks in the study of multi-modal population cohorts: an application to mental health," Université Paris-Saclay, 2024.

[10] S. Karim *et al.*, "Developments in brain tumor segmentation using MRI: deep learning insights and future perspectives," *IEEE Access*, vol. 12, pp. 26875–26896, 2024.

[11] C. Jin, T. F. Ng, and H. Ibrahim, "Advancements in Semi-Supervised Deep Learning for Brain Tumor Segmentation in MRI: A Literature Review," *AI*, vol. 6, no. 7, p. 153, 2025.

[12] T. Zhou, "M2GCNet: Multi-modal graph convolution network for precise brain tumor segmentation across multiple MRI sequences," *IEEE Trans. Image Process.*, 2024.

[13] B. Shaheema and N. B. Muppalaneni, "Multimodal brain image segmentation: a recent review, challenges and future perspectives," *Multimed. Tools Appl.*, pp. 1–48, 2025.

[14] R. R. Kumar and R. Priyadarshi, "Denoising and segmentation in medical image analysis: A comprehensive review on machine learning and deep learning approaches," *Multimed. Tools Appl.*, vol. 84, no. 12, pp. 10817–10875, 2025.

[15] Nguyen, T.-Q. T., et al. (2024). Brain Tumor Segmentation in MRI Images with 3D U-Net and Contextual Transformer. *arXiv:2407.08470*.

[16] A. Ali, R. Kumar, & S. Singh, *"Multimodal Semantic Segmentation Using Encoder-Based LinkNet Architecture for BraTS 2023 Challenge,"* IEEE international Conference on Biomedical Imaging, pp. 1–8, 2023.

[17] A. Sanaat, I. Shiri, S. Ferdowsi, H. Arabi, and H. Zaidi, "Robust-Deep: a method for increasing brain imaging datasets to improve deep learning models' performance and robustness," *J. Digit. Imaging*, vol. 35, no. 3, pp. 469–481, 2022.

[18] X. Liu *et al.*, "A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis," *lancet Digit. Heal.*, vol. 1, no. 6, pp. e271--e297, 2019.

[19] N. Prasad, S. W. A. Rizvi, and R. Saini, “Integrating multimodal imaging for precise brain tumour detection: a comprehensive image processing approach,” in *Recent Advances in Sciences, Engineering, Information Technology \& Management*, CRC Press, 2025, pp. 207–216.

[20] . Dehghani F, Karimian A, Arabi H. *Joint Brain Tumor Segmentation from Multi- imagnetic Resonance Sequences through a Deep Convolutional Neural Network.* Journal of Medical Signals & Sensors. 2024;14:9.

[21] Y. Ge, L. Xu, X. Wang, Y. Que, and M. J. Piran, “A novel framework for multimodal brain tumor detection with scarce labels,” *IEEE J. Biomed. Heal. Informatics*, 2024.

[22] M. Raghu, C. Zhang, J. Kleinberg, and S. Bengio, “Transfusion: Understanding transfer learning for medical imaging,” in *Advances in Neural Information Processing Systems*, 2019.

[23] P. H. Conze, G. Andrade-Miranda, V. K. Singh, V. Jaouen, and D. Visvikis, “Current and emerging trends in medical image segmentation with deep learning,” *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 7, no. 6, pp. 545–569, 2023.

[24] L. Blonde, K. Khunti, S. B. Harris, C. Meizinger, and N. S. Skolnik, “Interpretation and impact of real-world clinical data for the practicing clinician,” *Adv. Ther.*, vol. 35, no. 11, pp. 1763–1774, 2018.

[25] P. Saha, D. Mishra, F. Wagner, K. Kamnitsas, and J. A. Noble, “Examining modality incongruity in multimodal federated learning for medical vision and language-based disease detection,” *arXiv Prepr. arXiv2402.05294*, 2024.

[26] C. Huang, M. Jensen, S. Yeung-Levy, M. P. Lungren, H. Poon, and A. S. Chaudhari, “A Systematic Review and Implementation Guidelines of Multimodal Foundation Models in Medical Imaging,” *Res. Sq.*, Apr. 2025, doi:10.21203/rs.3.rs-5537908/v1.

[27] M. Jahanifar *et al.*, “Domain generalization in computational pathology: survey and guidelines,” *ACM Comput. Surv.*, vol. 57, no. 11, pp. 1–37, 2025.

[28] T. D. Pham, “Spatial uncertainty modeling of fuzzy information in images for pattern classification,” *PLoS One*, vol. 9, no. 8, p. e105075, 2014.

[29] M. L. Seghier, A. Ramlackhansingh, J. Crinion, A. P. Leff, and C. J. Price, “Lesion identification using unified segmentation-normalisation models and fuzzy clustering,” *Neuroimage*, vol. 41, no. 4, pp. 1253–1266, 2008.

[30] T. M. Quan, D. G. C. Hildebrand, and W. K. Jeong, “Fusionnet: A deep fully residual convolutional neural network for image segmentation in connectomics,” *Front. Comput. Sci.*, vol. 3, p. 613981, 2021.

[31] P. Saat, N. Nogovitsyn, M. Y. Hassan, M. A. Ganaie, R. Souza, and H. Hemmati, “A domain adaptation benchmark for T1-weighted brain magnetic resonance image segmentation,” *Front. Neuroinformatics*, vol. 16, p. 919779, Sep. 2022, doi: 10.3389/fninf.2022.919779.

[32] H. Zhang and Y. Qie, “Applying deep learning to medical imaging: a review,” *Appl. Sci.*, vol. 13, no. 18, p. 10521, 2023.

[33] B. Bonato, L. Nanni, and A. Bertoldo, “Advancing precision: A comprehensive review of MRI segmentation datasets from BraTS challenges (2012--2025),” *Sensors*, vol. 25, no. 6, p. 1838, 2025.

[34] F. R. Valverde, J. V Hurtado, and A. Valada, “There is more than meets the eye: Self-supervised multi-object detection and tracking with sound by distilling multimodal knowledge,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 11612–11621.

[35] X. Wang, “MMHA: A Multi-Modal Data Fusion Algorithm with Hierarchical Attention for Disease Diagnosis and Cancer Subtype Prediction,” *Expert Syst. Appl.*, p. 128205, 2025.