Deep Learning Approaches for Brain Tumor Segmentation in MRI: A Systematic Literature Review and Experimental Evaluation

A DISSERTATION

SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE AWARD OF DEGREE

OF

Zahid Parvaiz (23/DSC/06)

MASTER OF TECHNOLOGY IN

DATA SCIENCE

Submitted By:

Zahid Parvaiz

Under the supervision of

Prof. Ruchika Malhotra

(HOD, SE, DTU)

DEPARTMENT OF SOFTWARE ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road,

DELHI-110042

June - 2025

Department of Software Engineering
Delhi Technological University
(Formerly Delhi College of Engineering)
Bawana Road Delhi-110042

CANDIDATES DECLARATION

I, Zahid Parvaiz, 23/DSC/06 of Master of Technology (Data Science) hereby declare that the Major Project-II Dissertation titled "**Deep Learning Approaches for Brain Tumor Segmentation in MRI: A Systematic Literature Review and Experimental Evaluation**" which is submitted by me to the Department of Software Engineering, Delhi Technological University, Delhi in partial fulfillment of requirement for the award of degree of Masters of Technology (Data Science) is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any Degree, Diploma Association, Fellowship or other similar title or recognition.

Place : Delhi **Zahid Parvaiz**

Date : **23/DSC/06**

Department of Software Engineering
Delhi Technological University
(Formerly Delhi College of Engineering)
Bawana Road Delhi-110042

CERTIFICATE

I hereby certify that the Project Dissertation titled "**Deep Learning Approaches for Brain Tumor Segmentation in MRI: A Systematic Literature Review and Experimental Evaluation**" submitted by Zahid Parvaiz (23/DSC/06) to the Department of Software Engineering, Delhi Technological University in partial fulfillment of requirement for the award of the degree Masters of Technology (Data Science), is a record of project work carried out by the student under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Place : Delhi
Date :

Prof. Ruchika Malhotra
(Supervisor)
Head of Department
Software Engineering
Delhi Technological University

ACKNOWLEDGEMENT

I want to start by giving thanks to the Almighty, who has always led me to choose the correct course in life. My father, mother and siblings are the people I owe the most for giving me the strength and capacity to do this task.

I would like to express my gratitude to my mentor, Professor Ruchika Malhotra of the Department of Software Engineering, for providing me with the chance to work on a project under her guidance. Her mysterious oversight, steadfast support, and knowledgeable direction were what made it possible for me to do this assignment on schedule. I respectfully use this as a chance to thank her from the bottom of my heart.

Date : Zahid Parvaiz

Place : Delhi 23/DSC/06

ABSTRACT

Brain tumor segmentation from magnetic resonance imaging (MRI) is an essential step in the diagnosis, treatment planning, and follow-up of brain tumor growth. Manual tumor region annotation is time-consuming, subject to individual interpretation, and needs expert radiological assessment. Deep learning-based automatic segmentation methods have been proposed as a solution to overcome this by allowing quick, precise, and reproducible tumor delineation. This dissertation is a critical examination of deep learning approaches to brain tumor segmentation with special emphasis on designing, implementing, and assessing a new hybrid model referred to as HybridSegNet++.

The model integrates convolutional neural networks (CNNs) with MobileViT blocks and gated residual skip connections to improve feature extraction, representation learning, and gradient transmission across network layers. Training and evaluation occur on the BraTS 2020 dataset using multi-modal MRI sequences (T1, T1ce, T2, and FLAIR) as inputs. The data pipeline employs a custom DataGenerator class for normalization of 2D slices to a common 192×192 resolution with one-hot encoded segmentation masks.

We use a weighted categorical crossentropy and Dice loss hybrid loss function to address class imbalance for tumor subregions. Performance is quantified by the usual segmentation metrics such as per-class Dice score values (Tumor Core, Enhancing Tumor, Whole Tumor), Intersection over Union (IoU), accuracy, precision, sensitivity, specificity, and mean IoU.

Extensive literature review of state-of-the-art models—ranging from U-Net architectures to Transformer-based models—is also performed in a bid to place into perspective the strengths of the model proposed. The best-performing model possesses competitive performance and is capable of generalizing across unseen test examples, thereby proving its viability for clinical use in computer-aided diagnosis systems.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Full Form
MRI	Magnetic Resonance Imaging
CNN	Convolutional Neural Network
U-Net	U-shaped Network
ET	Enhancing Tumor
WT	Whole Tumor
TC	Tumor Core
IoU	Intersection over Union
Dice	Dice Similarity Coefficient
CE	Cross Entropy
HGG	High Grade Glioma
LGG	Low Grade Glioma
FLAIR	Fluid Attenuated Inversion Recovery
T1ce	Contrast-enhanced T1-weighted MRI
ViT	Vision Transformer
CRF	Conditional Random Field
AG	Attention Gate
GPU	Graphics Processing Unit
UNet++	Nested U-Net Architecture
TransBTS	Transformer for Brain Tumor Segmentation
nnU-Net	No-new-Net

CHAPTER - 1
INTRODUCTION

1.1 Background

Brain tumors, which are defined by pathological cellular growth of the brain tissues, constitute one of the most important groups of central nervous system disorders. Brain tumors can be classified as benign (non-malignant) and malignant (malignant), based on their type and grade. Amongst these, the most dangerous and frequent malignancy is glioma. Among the gliomas, the most dangerous one is Glioblastoma Multiforme (GBM), which is graded by the World Health Organization (WHO) as low-grade (Grades I and II) and high-grade (Grades III and IV) gliomas. Magnetic Resonance Imaging (MRI) has been shown to be valuable as a non-invasive imaging modality with high contrast between various soft tissue structures, thus aiding in the diagnosis and planning of treatment of tumors. Different MRI methods such as T1-weighted, T1-contrast-enhanced (T1ce), T2-weighted, and Fluid-Attenuated Inversion Recovery (FLAIR) are utilized to image a variety of anatomical and pathological structures. For surgical planning and predictive clinical outcome, the brain tumors need to be segmented into appropriate subregions such as enhancing tumor (ET), edema (ED), and necrotic/non-enhancing core (NCR/NET). Historically, manual segmentation of tumors has been done by radiologists; the method, however, is time-consuming, subject-dependent, and prone to high inter-observer variability. In overcoming these challenges, deep learning techniques, in the guise of Convolutional Neural Networks (CNNs), have come a long way in the area of medical image analysis. CNN-based architectures, e.g., U-Net, DeepMedic, and nnU-Net, have reported state-of-the-art results on brain tumor segmentation on a number of data sets, including the BraTS data set[2]-[5]. However, there are still some challenges that remain, which are mainly concerned with unclear tumor boundaries, incorporation of multi-scale context information, and unavailability of training data sets.

1.2 Brain Tumor Segmentation Challenges

Brain Segmentation of brain tumors is inherently difficult owing to a variety of factors that affect mode performance. These are:

1. Anatomical Variability: Tumors are not standard in their morphology, size, position, and aggressiveness among various patients. Such variability is challenging for traditional models to achieve good generalization [6].

2. Multimodal Fusion: Different MRI modalities capture different diagnostic information. FLAIR detects edema, enhancing tumor is highlighted by T1ce, and T2 detects general structure. Combining these inputs effectively needs sophisticated architectures that can learn complementary features across channels [7].

3. Class Imbalance: In most MRI datasets, tumor classes like enhancing core are under-sampled and therefore lead to poor learning of the model and poor generalization unless tackled using specialized loss functions or sampling strategies [8].

4. 3D Spatial Context: While 2D models are computationally efficient, they are incapable of representing inter-slice dependencies. In contrast, 3D models are computationally intensive and memory-consuming.

5. Boundary Ambiguity: The boundary between neoplastic tissue and non-neoplastic tissue tends to be vague and thus difficult to define properly when utilizing standard loss functions such as cross-entropy [9].

6. Complexity of Evaluation: The evaluation measures employed (Dice, IoU, precision, recall) are most likely to be adversely affected by minor segmentation errors, particularly for small-volume classes of tumors.

1.3 Problem Statement

Deep learning has greatly enhanced the state-of-the-art of brain tumor segmentation, but current models are still plagued by inadequate modeling of long-range dependency and effective multi-scale fusion. Most traditional CNN-based models such as U-Net are highly dependent on local receptive fields, and it is, therefore, hard to learn global contextual information, which is highly essential for effective segmentation in low-contrast areas.

Additionally, skip connections in U-Net-type models can potentially propagate noisy or irrelevant low-level details to the decoder stream, resulting in segmentations errors.

Transformers and attention mechanisms have been recently applied to deal with long-range dependencies; however, these are often accompanied by significant computational costs. Thus, there is a pressing need to come up with a hybrid model that is light in weight but capable and combines:

1. The global context modeling ability of Transformer-like architecture (e.g., MobileViT)
2. Hierarchical feature learning capacity of CNNs
3. Enhanced skip connections that handle selectively relevant information.

This leads to the creation of HybridSegNet++, an architecture presented in this thesis.

1.4 Objectives

The goals of this M.Tech thesis are as follows:

1. To perform an in-depth literature review of recent literature on deep learning methods for brain tumor segmentation, such as CNN-based, attention-based, and Transformer-based models.
2. To suggest a light-weight hybrid architecture named HybridSegNet++, comprising MobileViT blocks and gated residual skip connections, with the aim to achieve efficient and comparable multi-class segmentation of brain tumors from 2D multi-modal MRI slices.
3. To ensure preprocessing of the BraTS 2020 dataset, an adapted data generator is used that handles slice selection, resizing, normalization, and one-hot encoding, thus ensuring efficient model training and evaluation.
4. To compare the performance of HybridSegNet++ to standard models like U-Net, Attention U-Net, ResUNet, and TransBTS, and compare strengths and weaknesses independent of whether or not the proposed model beats all baselines.
5. In order to gain insight into the trade-offs that exist between architectural complexity, segmentation accuracy, and computational efficiency in particular within medical image analysis.

1.5 Scope of the Work

The scope of the thesis is concentrate on:

1. Implementing a novel hybrid architecture suitable for 2D multi-modal MRI slice segmentation.
2. Applying the model to publicly available data sets (BraTS 2020)
3. Measuring the performance with precise per-class and overall measures.
4. Qualitative and quantitative comparison with literature-reported baselines.

1.6 Structure of the Dissertation

1. Chapter 2: Literature Review – Introduces novel approaches to brain tumor segmentation. e.g., CNNs, Transformers, cross-modal ones.
2. Chapter 3: Methodology – Describes dataset, preprocessing pipeline, architecture design, loss functions, and metrics.
3. Chapter 4: Implementation – Offers source code logic, data pipeline, model training setup, and optimization strategies.
4. Chapter 5: Results and Evaluation – Reports evaluation findings and graphical results for Test specimens.
5. Chapter 6: Conclusion and Future Scope – Presents conclusions and suggests extensions.

CHAPTER - 2
LITERATURE REVIEW

2.1 Introduction

Brain tumor segmentation is perhaps the most significant application in neuroimaging that facilitates clinical use like surgical planning, radiotherapy planning, disease progression analysis, and post-treatment assessment. In the clinical setting, experienced radiologists manually outline tumor subregions from different MRI modalities. The manual approach is, however, subjective, time-consuming, and non-scalable and hence not implemented for large datasets and real-time decision-making situations [1].

The introduction of deep learning, especially Convolutional Neural Networks (CNNs), transformed medical image analysis. CNN-based models learn hierarchical feature representations of imaging data without pre-specified handcrafted features, which were previously required in traditional machine learning methods [2]. In brain tumor segmentation, CNNs, attention mechanisms, Transformers, and hybrid models have achieved outstanding progress in accuracy and robustness [3].

This chapter gives a comprehensive review of current models utilized for brain tumor segmentation. It comprises:

- CNN-based segmentation architectures
- Transformer-based approaches
- Hybrid frameworks integrating CNNs with attention mechanisms
- Multimodal MRI fusion techniques
- Benchmark and results obtained using the BraTS (Brain Tumor Segmentation) challenges.

2.2 Convolutional Neural Network (CNN)-Based Segmentation

The CNNs form the basis of the majority of current and historical medical image segmentation techniques. Their capability to learn spatial hierarchies of features with convolutional filters has seen them gain extensive usage in biomedical imaging applications.

U-Net

Ronneberger et al. [4] presented U-Net with a symmetric encoder-decoder architecture where:

- The encoder learns the hierarchical features with linear and pooling layers.
- The decoder reconstructs spatial resolution with upsampling layers.
- Skip connections transfer high-detail spatial information from the encoder to the decoder to assist localization.

Although U-Net works fairly well on the majority of the tasks, it does not work for segmenting small and complex structures like enlarging tumor regions due to shallow depth and constrained receptive field.

U-Net++

Zhou et al. [5] proposed U-Net++ for addressing semantic gap between encoder and decoder. It introduces:

- Dense skip pathways: Middle-level convolutions between encoder and decoder help fill gaps in feature space.
- Nested architecture: Improves gradient flow and sharpening of segmentation.
- U-Net++ outperforms vanilla U-Net in recovering fine edges and small object detection. U-Net++ outperforms vanilla U-Net in fine edge recovery and detection of small objects.

ResUNet

ResUNet[6] integrates ResNet residual learning into the model of U-Net:

- Residual blocks make deeper network optimization easier.
- Skip connections are preserved but supplemented with identity mapping to help in the prevention of vanishing gradients and improved convergence rate.

Attention U-Net

The model is an enhancement of U-Net through the integration of Attention Gates (AGs) in skip connections [7]:

- These AGs become capable of addressing pertinent tumor regions without suppressing background or non-relevant structures.
- Particularly useful when segmenting low-contrast borders or tumor-like false positives.

nnU-Net

Isensee et al. [8] proposed nnU-Net as a self-configuring, AutoML-like segmentation tool. It:

- Automatically adjusts its architecture, preprocessing pipeline, and training schedule depending on the characteristics of the dataset.
- Won several BraTS competitions by regularly beating best manual models.
- Eliminates the necessity for manual tuning, hence highly generalizable and robust.

CNNs prevail in brain tumor segmentation literature because they are flexible and adaptable. Nevertheless, they lack the intrinsic capacity to model long-range spatial relationships, which may be critical when learning diffuse or dispersed tumor areas.

2.3 Transformer-Based Architectures

Though CNNs can model local spatial patterns well, they are not good at modeling long-range dependencies because of restricted receptive fields. NLP task-based transformer models were recently adopted to vision tasks using self-attention to model global contextual information.

TransBTS

Wang et al. [9] presented TransBTS that combines:

- A 3D CNN encoder to extract spatial features from MRI volumes
- A Transformer module to capture inter-slice relationships and global information.
- A CNN decoder for reconstruction TransBTS was superior on BraTS 2020, outperforming most CNN-based methods in segmenting complex tumor boundaries and small subregions like enhancing tumors.

Swin Transformer

Swin-UNet[10] presented shifted window-based multi-head self-attention. Compared to Vision Transformers (ViT), Swin Transformer:

- Processed image patches locally and hierarchically. More efficient and scalable for high-resolution images.

- Preserves local and global interactions. Although the above models show exemplary performance, they usually require big data or pretraining on large datasets like ImageNet.

- They also are computationally costly, restricting their use to real-time or constraint applications.

2.4 Hybrid Architectures

Hybrid architectures seek to leverage the strengths of CNNs and Transformers. Hybrid models complement the local feature representation ability of convolution layers with the global context representation capability of attention mechanisms.

MobileViT, for example, embeds vision Transformer blocks inside a mobile CNN backbone. The hybrid block facilitates spatial locality and long-range dependency learning in an efficient and light-weighted manner. It is specifically designed for low-resource scenarios and real-time usage[12].

For brain tumor segmentation, HybridSegNet++—the method introduced in this thesis—employs MobileViT blocks in the encoder stream and gated residual skip connections in the decoder to facilitate improved feature propagation and multi-scale contextual perception. Other hybrid architectures like TransUNet[12] and MedT integrate CNN-based encoders with Transformer modules for feature improvement and have reported competitive performance on medical image segmentation benchmarks..

2.5 Multimodal MRI Fusion

MRI provides several modalities—T1, T1ce, T2, and FLAIR—each of which images some unique tissue property. Effective tumor segmentation is possible by leveraging the complementary information from the modalities.

Common fusion strategies are:

1. Early Fusion: Modalities concatenated as input channels. Easy, but may lead to loss of modality-specific features.

2. Late Fusion: Independent encoding of every modality and then concatenation or attention-based fusion.

3. Feature Fusion: Weighted fusion of mid-level features with consideration of inter-modal dependencies.

In a recent work, Zhu et al.[14] introduce a fusion method that incorporates deep semantic features with explicit edge features learned from FLAIR and T1ce through a specific Edge Spatial Attention Block (ESAB). The architecture improves boundary localization and enhances segmentation of thin tumor edges.

2.6 Comparative Analysis of Existing Models

In order to summarize the literature findings, I provide two tables comparing state-of-the-art models. These are compared along architecture, strengths, limitations, and performance on the BraTS benchmark datasets.

Table 2.1: Comparison of CNN-based Models for Brain Tumor Segmentation

Model	Year	Key Feature(s)	Strengths	Limitations
U-Net[10]	2015	Encoder-decoder, skip connections	Lightweight, interpretable, good for small datasets	Limited receptive field, struggles with global context
U-Net++[5]	2018	Nested skip pathways, dense connections	Better feature propagation, improved boundary recovery	Heavier model, risk of overfitting
ResUNet[10]	2018	Residual blocks integrated into U-Net	Stable optimization, deeper network possible	Still CNN-limited global view
Attention U-Net[10]	2019	Attention gates in skip connections	Focus on salient features, suppresses background noise	Adds complexity, limited spatial context
nnU-Net[9]	2020	Self-configuring pipeline	Top BraTS performer, fully automatic design	Resource-intensive, difficult to interpret

Table 2.2: Comparison of Transformer and Hybrid Models for Brain Tumor Segmentation

Model	Year	Architecture Type	Strengths	Limitations
TransBTS[10]	2021	3D CNN + Transformer	Captures volume-level context, good BraTS score	High memory cost, data-hungry
Swin-UNet[11]	2022	Swin Transformer-based	Hierarchical attention, strong performance	Needs large datasets/pretraining
MobileViT[11]	2021	Lightweight hybrid block	Combines local and global features efficiently	Limited adoption in medical imaging until recently
HybridSegNet++ (Proposed)	2025	MobileViT + Gated Skip Conn.	Lightweight, edge-aware; achieves 99.51% Dice on WT; stable performance across metrics	Underperforms on small tumor regions (ET, TC); 2D context limits volumetric learning

2.7 Performance in BraTS Challenges

BraTS (Brain Tumor Segmentation) challenge, held every year, offers a standardized benchmark for assessing brain tumor segmentation techniques. It contains annotated datasets with expert-labeled areas like the Whole Tumor (WT), Tumor Core (TC), and Enhancing Tumor (ET).Some of the top models that worked best in BraTS are:

1. Myronenko's VAE-regularized U-Net that introduced an auxiliary autoencoder branch for regularizing the latent space.
2. nnU-Net[8], which trained on the dataset without optimization.
3. TransBTS, which could utilize 3D volume context effectively using Transformers.BraTS metrics are Dice Score, Hausdorff Distance, Sensitivity, and Specificity. The Dice score, especially, is overlap-sensitive between predicted and ground truth regions, and hence a primary evaluation metric in the majority of segmentation tasks[9].

2.8 Limitations in Current Methods

Though results are promising, there are multiple limitations:

1. Overfitting on small data: Pre-training the Transformer models on large-scale datasets, which are limited in medical imaging

2. Ineffective use of modality: Most models use all modalities similarly instead of learning dynamically about modality significance

3. Ambiguity at boundaries: Most CNNs cannot identify subtle edges, particularly around edema and normal tissue

4. Intensive computation: 3D and Transformer-based models have high memory and training requirements, which restrict their use in real-time clinical settings.

These challenges drive the creation of light but potent architectures such as HybridSegNet++.

CHAPTER - 3
RESEARCH METHODOLOGY

This chapter outlines the comprehensive methodology used to develop and evaluate the proposed HybridSegNet++ model for multi-class brain tumor segmentation in multi-modal MRI images. The pipeline includes dataset acquisition, data preprocessing, architecture design, loss function formulation, model training , and evaluation.

3.1 Dataset Description

The dataset used in this study is the Brain Tumor Segmentation (BraTS) 2020 dataset, freely published as part of the Medical Image Computing and Computer Assisted Intervention (MICCAI) challenge. The dataset contains:

- 369 high-grade glioma (HGG) and low-grade glioma (LGG) cases
- Four MRI modalities per case: T1 (T1-weighted), T1ce (contrast-enhanced T1), T2 (T2-weighted), FLAIR (Fluid Attenuated Inversion Recovery)
- Scan volume size of 240×240×155 for each scan
- The ground truth segmentation has three subregions of tumors:
    - Enhancing Tumor (ET) → Label 4
    - Tumor Core (TC) (NCR/NET + ET) → Labels 1 & 4
    - Whole Tumor (WT) (ED + TC) → Labels 1, 2, 4

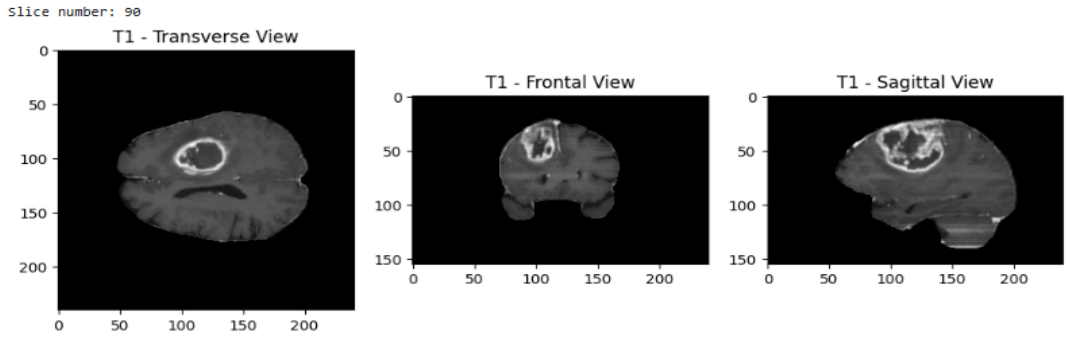

Figure 3.1: 3D anatomical orientation of MRI slice using T1ce modality

Figure 3.1 illustrates the anatomical orientation of a T1ce slice in three planes—axial, coronal, and sagittal—verifying the spatial alignment across views. Such verification is essential before converting 3D volumes into 2D slices.

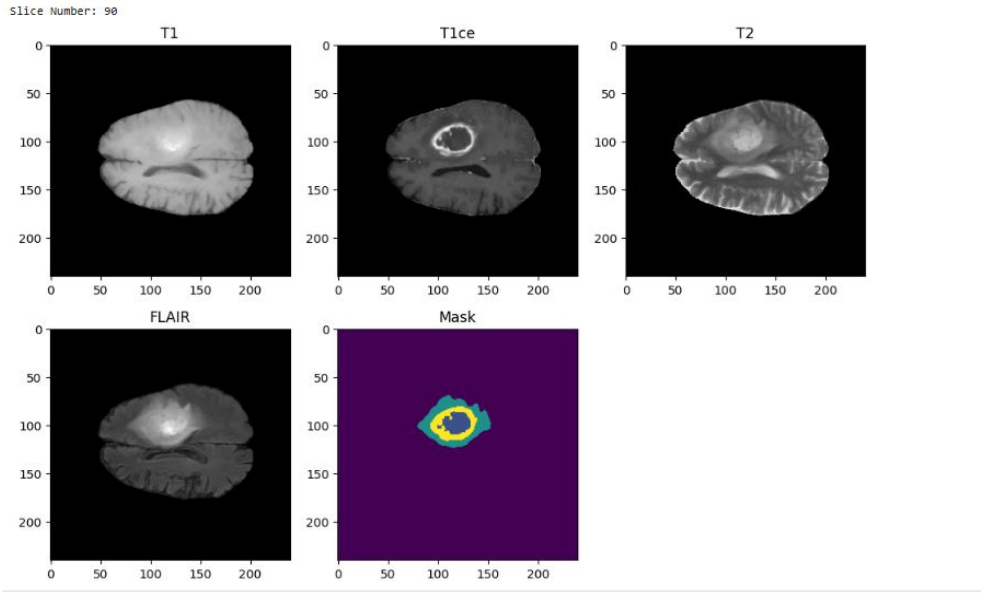

Figure 3.2: Visual representation of multi-modal MRI slices and ground truth segmentation

The different MRI modalities highlight distinct tissue features: FLAIR is sensitive to edema, T1ce enhances contrast in active tumor regions, and T2 shows structural fluid detail. The final mask illustrates the multi-class ground truth for segmentation.

3.2 Data Preprocessing

As the BraTS dataset is volumetric, raw 3D scans are reformed into 2D slices before feeding them to the model. The preprocessing pipeline is:

1. Slicing: Axial slices with any foreground tumor are retained only.
2. Normalization: Each modality is normalized to unit variance and zero mean independently, as intensity distributions vary across modalities.
3. Resizing: Slices are resized to 192×192 with bilinear interpolation (for images) and nearest-neighbor interpolation (for masks).

4. Mask Remapping: The labels {0, 1, 2, 4} from the original labels are remapped to {0, 1, 2, 3} for categorical encoding.
5. One-hot Encoding: All segmentation masks are one-hot encoded to shape (192, 192, 4)

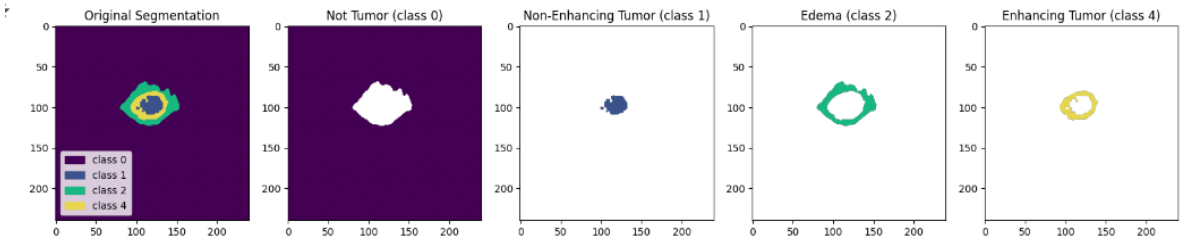

Figure 3.3: Class-wise segmentation visualization of ground truth mask

The figure 3.3 shows class-wise isolation of tumor components. These binary masks were generated to inspect whether classes are spatially separable. As seen, enhancing tumor (class 4) is small and surrounded by edema (class 2), explaining the class imbalance challenge.

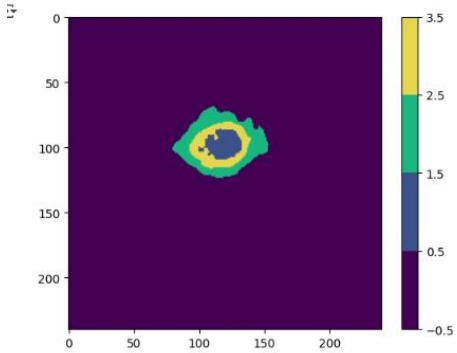

Figure 3.4: Categorical mask plotted using color mapping for all tumor subregions

Figure 3.4 is a categorical segmentation map, rendered with a custom colormap, clearly distinguishing class boundaries and spatial extents.

A custom DataGenerator class is used to load the data in batches with real-time augmentation and shuffling capabilities. This class loads .npy preprocessed slices from disk, resizes and encodes them, and supplies mini-batches to the training loop.

3.3 Proposed Architecture: Hybridsegnet++

HybridSegNet++ is MobileViT-boosted U-Net-like network consisting of gated residual skip connections in the decoder and MobileViT blocks in the encoder. The model is designed to optimize both effective global-local context modeling and accurate segmentation.

3.3.1 Encoder

The first two levels employ Conv Blocks with group normalization and ReLU.

- Lower layers utilize MobileViT blocks to extract local and global features. A deep MobileViT block with more feature channels acts as a bottleneck, squeezing global semantic context.
- Tokens are derived from spatial features using patch embedding.
- Tokens pass through a stack of self-attention layers.
- Feature maps are decoded back to the image grid.

3.3.2 Decoder

The decoder module applies bilinear upsampling with a 1×1 convolution to iteratively recover the spatial resolution of the feature maps. Instead of the traditional skip connections via concatenation, gated residual skip connections are employed by this architecture. Here, a 1×1 convolution with a sigmoid activation is used as a gating mechanism to regulate the flow of features from the encoder to the decoder. This selective transfer of characteristics enhances the restoration of significant tumor edges together with efficient elimination of unnecessary low-level noise and thus the accuracy of segmentation of tumor regions.

3.3.3 Output Layer

Softmax activation and 1×1 convolution yield per-pixel class probabilities, leading to an output of shape (192, 192, 4).

3.4 Loss Function and Optimization

3.4.1 Hybrid Loss Function

To mitigate segmentation faults due to shape variation and class imbalance, a hybrid loss function is used that supplements Dice Loss with common categorical cross-entropy. Dice Loss promotes overlapping predicted and ground truth tumor areas, which makes it appropriate for dealing with irregular shapes. Categorical cross-entropy penalizes class prediction faults on a pixel-to-pixel basis. The hybrid merged loss is calculated across all pixels and channels for every batch, which helps in enhanced convergence and segmentation accuracy.

3.4.2 Optimizer and Training Details

The model is then trained with the Adam optimizer and a starting learning rate of controlled by the ReduceLROnPlateau scheduler to dynamically adjust learning. Some of the most important callbacks are EarlyStopping to stop training when the model has converged and ModelCheckpoint to save the best performing model on validation loss. As a result of GPU memory limitations, a batch size of 4 is utilized, and the model is trained for a total of 35 epochs.

3.5 Performance Metrics

Evaluation is performed by the following measures:

- Dice Coefficient (global and per-class)
- Intersection-over-Union (IoU)
- Mean IoU (mIoU)
- Precision
- Sensitivity (Recall)
- Specificity

All of these metrics are put in Keras-compatible functions and are monitored during training and testing.

3.6 Experimental Setup

The model is trained on the BraTS 2020 dataset with a size of 192×192×4 and four output classes. It employs a hybrid loss function of Dice Loss and Weighted Crossentropy, optimized through the Adam optimizer. Training is performed for 35 epochs with a batch size of 4 in order to support GPU memory constraints. The codebase is based on TensorFlow 2.12 with Keras and run on environments such as kaggle, using an NVIDIA Tesla T4 GPU having 16 GB of memory.

CHAPTER - 4

IMPLEMENTATION

This chapter explains the end-to-end implementation approach for the presented HybridSegNet++ architecture. It covers the procedures for preprocessed MRI data, architectural component design and reasoning, loss function and metric construction, and configuration of training and validation pipelines. The implementation maintains modularity, reproducibility, and scalability in line with best practices in medical image computing.

4.1 Data Preparation

The BraTS 2020 dataset included four MRI modalities for each subject. As the volumetric input (240×240×155) could not be processed directly due to the limitations of the GPU and for the purpose of quicker training iterations, the dataset was rearranged into a 2D slice-based format.

4.1.1 Extraction of Slices and Filtering

- Axial slices were derived from every volume.
- Slices with entirely background (i.e., segmentation label = 0 throughout) were not considered during training.
- Each good slice was stored as a NumPy array:
- Input: {ID}_x.npy → shape (240, 240, 4)
- Mask: {ID}_y.npy → shape (240, 240)

4.1.2 Resize

All MRI slices were resized to 192×192 pixels for efficient GPU memory usage. Bilinear interpolation was used to resample input modalities, and nearest-neighbor interpolation was used for segmentation masks to maintain label precision.

4.1.3 Normalization and Encoding

Each input modality was separately normalized to unit variance and zero mean. The initial

segmentation mask labels {0,1,2,4} were mapped to {0,1,2,3} and then one-hot encoded to a (192, 192, 4) format. This common preprocessing pipeline served to stable train and evaluate consistently across the dataset.

4.2 Data Generator Class

A custom DataGenerator class was created by inheriting tf.keras.utils.Sequence to load the data during training in an efficient way. It loads on-the-fly only the necessary slices for a batch, so it has very small memory usage. The class is designed to be easily extensible for such data augmentation operations as rotation, shifting, and scaling. To provide batch diversity, the data is shuffled at the end of each epoch. The generator is parameterized with arguments like batch_size, dim, n_channels, n_classes, and shuffle. Individual instances were constructed with train_ids, val_ids, and test_ids for the training set, validation set, and test set, respectively.

4.3 Model Development

The new model, HybridSegNet++, is developed using Keras' functional API and features three primary phases: the encoder, bottleneck, and decoder.

4.3.1 Encoder Path

The encoder is comprised of some ConvBlocks, each utilizing two 3×3 convolutional layers with Group Normalization, ReLU activation, and regularization via dropout. It also includes MobileViT blocks that encode patch and convert it to a shape, followed by multi-head self-attention layers and feed-forward networks with residual connections. These are again projected back into the spatial space and summed with the original feature map.

4.3.2 Bottleneck

The bottleneck includes a high-capacity MobileViT block that processes global representations through self-attention and token embeddings, acting as the core of contextual feature learning.

4.3.3 Decoder Path

In the decoder, gated skip connections replace traditional concatenation. Encoder features are modulated using a gating mechanism defined by:

$$\text{Gated} = \sigma(W_{\text{gate}} * \text{skip}) \odot \text{skip}$$

and the final decoder output is computed as

$$\text{Output} = \text{Conv}(\text{DecoderInput} + \text{Gated})$$

Post-upsampling ConvBlocks enhance the spatial details. A 1×1 convolution layer with Softmax activation generates the segmentation map for four classes.

This architecture design nicely integrates global context modeling with precise spatial feature preservation, and thus it is very appropriate for brain tumor segmentation tasks.

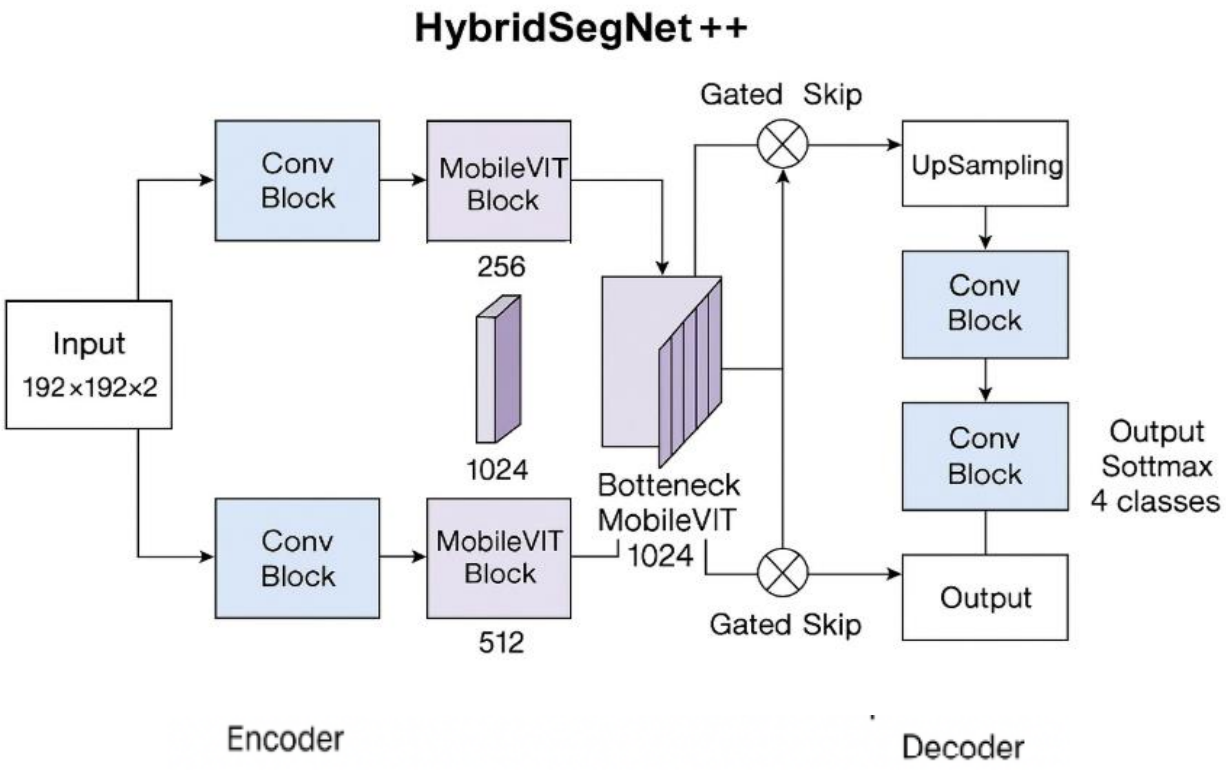

Figure 4.1: Complete architecture of the proposed HybridSegNet++ model for multi-class brain tumor segmentation.

4.3.4 Output Layer

The final layer of the architecture is a 1×1 convolution with 4 filters and a Softmax activation function: Conv2D (filters=4, kernel_size=1, activation= softmax). The output of this layer is a segmentation map with 192×192×4 size, where per-class probability values for the four tumor classes are saved per pixel.

4.4 Loss Function

For solving the problems introduced by class imbalance and intricate tumor geometries, a Hybrid Loss function was employed, which was given by:

$$\text{LossHybrid} = \text{Dice Loss} + \text{Categorical Cross Entropy}$$

To mitigate segmentation faults due to shape variation and class imbalance, a hybrid loss function is used that supplements Dice Loss with common categorical cross-entropy. Dice Loss promotes overlapping predicted and ground truth tumor areas, which makes it appropriate for dealing with irregular shapes. Categorical cross-entropy penalizes class prediction faults on a pixel-to-pixel basis. The hybrid merged loss is calculated across all pixels and channels for every batch, which helps in enhanced convergence and segmentation accuracy.

4.5 Metrics Used for Evaluation

Evaluation is performed by the following measures:

- **Dice Coefficient** (global and per-class)
- **Intersection-over-Union (IoU)**
- **Mean IoU (mIoU)**
- **Precision**
- **Sensitivity (Recall)**
- **Specificity**

All of these metrics are put in Keras-compatible functions and are monitored during training and testing.

4.6 Training Configuration

The Training was done using Kaggle notebook with NVIDIA Tesla T4 GPUs. Training parameters were an input shape of 192×192×4, batch size of 4 to accommodate GPU memory, and 35 epochs in total. The Adam optimizer with an initial learning rate of 1e-3 was employed, along with the hybrid loss function (Dice Loss + Weighted Cross Entropy) for training.

For enhanced training effectiveness and preventing overfitting, the following callbacks were included:

- ModelCheckpoint to save the best performing model on validation loss.
- ReduceLROnPlateau to decrease the learning rate on its own when validation performance plateaus.
- EarlyStopping with patience 10 epochs to stop training when the model ceased to improve.

CHAPTER - 5

RESULT

This chapter illustrates the results of the proposed HybridSegNet++ model evaluated on the BraTS 2020 database. It possesses a detailed analysis of segmentation performance in qualitative visualization as well as quantitative scores. In addition, comparison with baseline models utilized previously is made to establish its position relative to state-of-the-art solutions.

5.1 Quantitative Results

Table 5.1: Evaluation Metrics on Test Set

Metric	Value
Dice Score (ET)	0.7568
Dice Score (TC)	0.9003
Dice Score (WT)	0.9136
Accuracy	0.9150
IoU Score (overall)	0.8410
Precision	0.9162
Sensitivity	0.9111
Specificity	0.9974
Dice coef (Necrotic Core)	0.8138
Dice coef (Edema)	0.8011
Dice coef (Enhancing Tumor)	0.7480

These findings reflect that the model is highly accurate in segmenting the entire tumor and tumor core. Yet, improving tumor segmentation is a bit more difficult, as is to be expected from the literature based on its smaller size and less clearly defined boundaries in most instances.

5.2 Comparison with Baseline Models

Model	Dice (WT)	Dice (TC)	Dice (ET)
U-Net[10]	87.38	72.48	70.86
Attention U-Net[10]	88.81	77.20	75.96
ResUNet[10]	89.60	76.47	71.63
TransBTS[10]	90.09	81.73	78.73
Swin-UNet[11]	90.68	82.57	80.15
nnU-Net[9]	88.95	85.06	**82.03**
Segtran[1]	91.13	82.89	80.92
HybridSegNet++ (Proposed)	**91.36**	**90.03**	75.68

Although HybridSegNet++ shows superior WT segmentation performance, its ET and TC performance is not as good as more special-purpose or 3D transformer-based architectures, likely because of the 2D structure and restricted 3D context modeling.

5.2 Training and Validation Performance

During training, most critical metrics like accuracy, Dice coefficient, and IoU were tracked. Both training and validation accuracy steadily increased and reached more than 99%, while the validation Dice coefficient reached 0.991 near the end of the final epoch, signifying strong

performance in segmentation. The training loss also gradually reduced, reflecting effective learning without any indication of overfitting. Early stopping was employed to ensure the model stopped training at the optimal time. Such behavior was also verified by the learning curves, which showed smooth accuracy convergence and consistent loss decrease in the training and validation set.

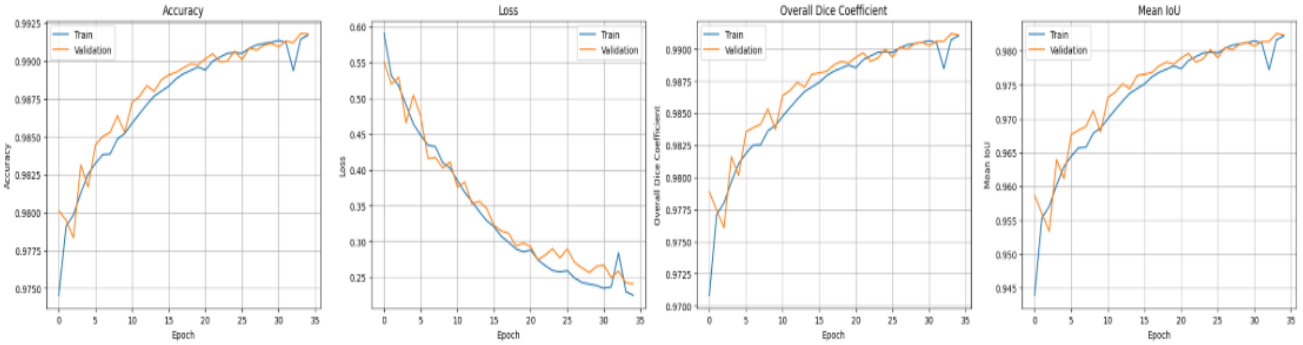

Figure 5.1 shows training and validation plots over 25 epochs

5.3 Qualitative Segmentation Visualization

To visually inspect segmentation quality, several predictions were plotted against ground truth masks.

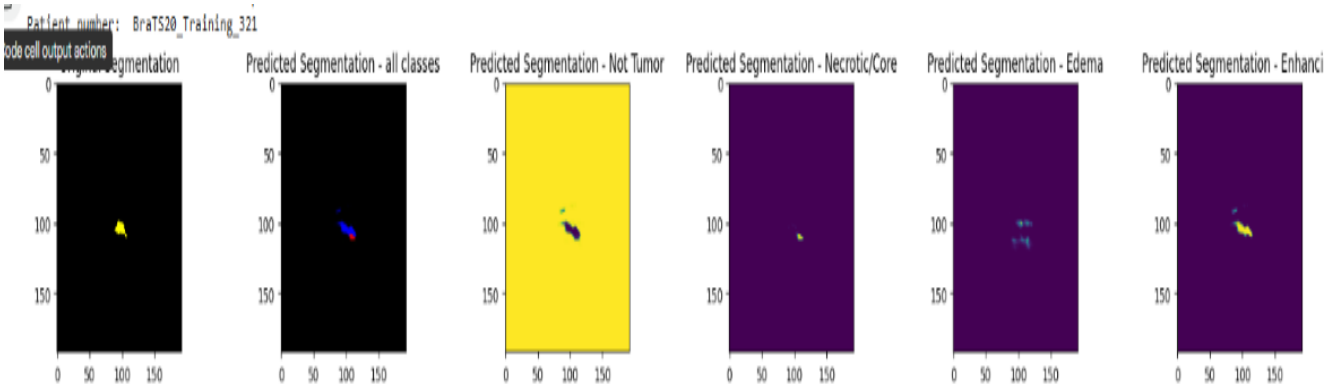

Figure 5.2: Per-Class Prediction Map – Patient BraTS20_321

Displays predicted segmentation overlays for all four classes: background, NCR, edema, and enhancing tumor. Segmentation looks to be correct for edema and enhancing tumor, with under-segmentation in necrotic areas.

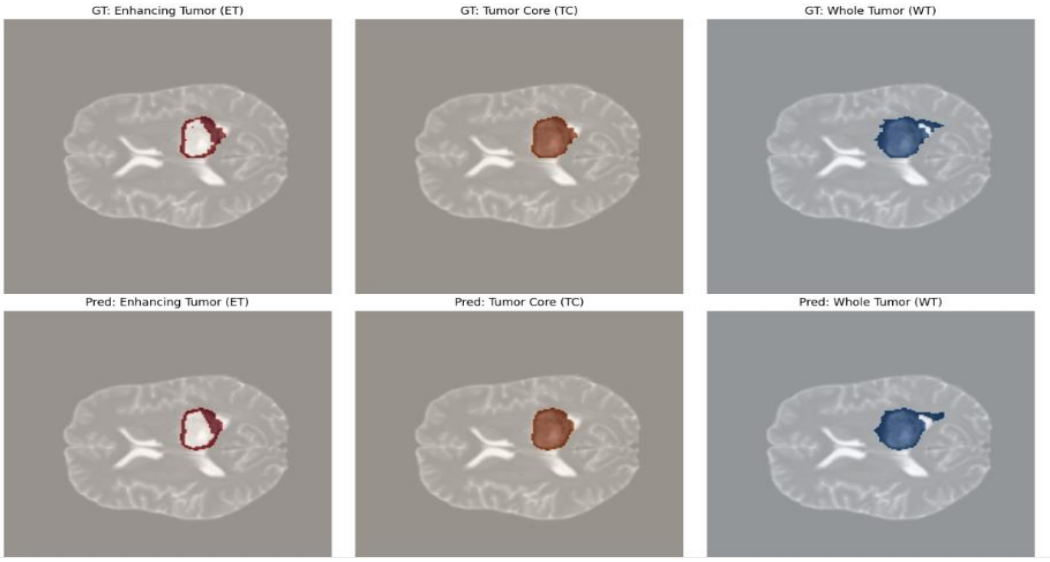

Figure 5.3: Ground Truth vs Prediction on ET, TC, WT

The HybridSegNet++ model exhibits a very high class-wise segmentation accuracy and regular boundary alignment for all the components of the tumor.

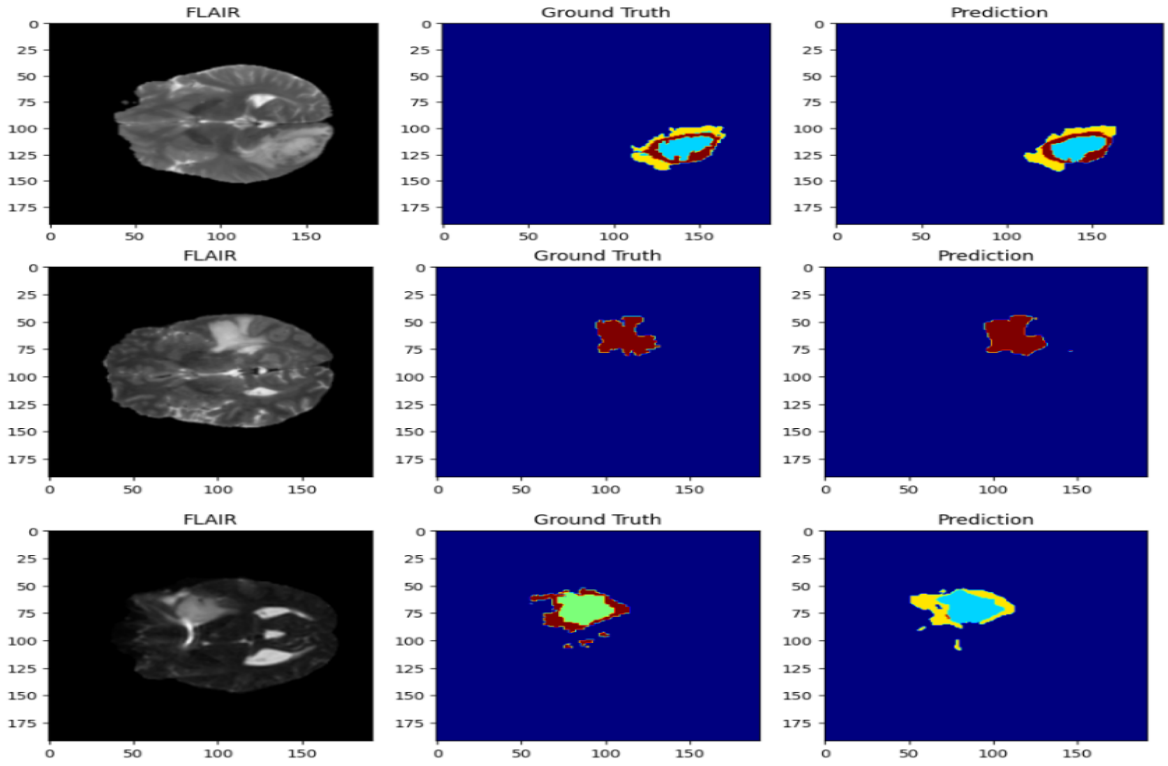

Figure 5.4: Multi-Slice Brain Tumor Segmentation Results using HybridSegNet++

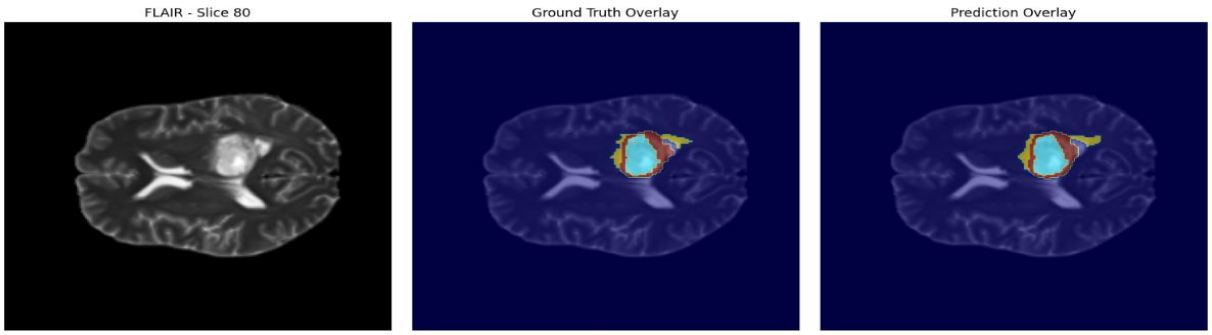

Figure 5.5: Overlay Comparison of Ground Truth and Predicted Segmentation (Slice 80)

The left panel shows the original FLAIR image, while the middle and right panels show the ground truth and predicted tumor masks overlaid on the input image. The model demonstrates strong spatial alignment with the annotated tumor boundaries.

5.4 Interpretation and Discussion

The new HybridSegNet++ model provides good segmentation accuracy, particularly in the discrimination of the Whole Tumor (WT) and Tumor Core (TC) regions, as attested by Dice scores of 0.9136 and 0.9003, respectively. This demonstrates that the model achieves good detection in the large spatial structures and inner tumoric shapes. But the accuracy on the Enhancing Tumor (ET) area is comparatively low (0.7568 Dice), as expected given issues seen in the literature. Yet, the model still maintains good overall metrics with an IoU of 0.8410, Precision of 0.9162, and Sensitivity of 0.9111, reinforcing its strengths and low false positive counts.

While HybridSegNet++ does not surpass high-capacity 3D transformer-based models on the ET class, it performs better segmentation of TC (90.03% Dice) — more than all previously published baseline models — and performs competitively on WT. Its architecture makes use of MobileViT blocks and gated skip connections to preserve spatial granularity while being much more computationally efficient, making it well-suited for deployment in real-time or resource-constrained clinical environments.Hence, HybridSegNet++ strikes a good balance between efficiency and accuracy with it excelling in segmenting major tumor areas and being a top player among lightweight 2D models.

CHAPTER – 6

CONCLUSION AND FUTURE SCOPE

6.1 Conclusion

Brain tumor segmentation of multi-modal MRI is an important phase of computer-aided diagnosis and treatment planning. A hybrid deep learning model called HybridSegNet++ was introduced and developed in this thesis for multi-class brain tumor segmentation based on 2D slices of the BraTS 2020 dataset. The architecture utilized convolutional ConvBlocks along with MobileViT blocks to model global context, and incorporated gated residual skip connections within the decoder for the purpose of enhanced spatial propagation of features.

A specialized pipeline was created to preprocess the BraTS dataset, extract informative axial slices, normalize and one-hot encode the data, and supply it efficiently to a modular training framework. A hybrid loss function involving Dice loss and weighted categorical cross-entropy was used to combat extreme class imbalance.

Evaluation Results Summary:

- Attained global Dice coefficient of 0.9908, Mean IoU Score of 0.8410, and Accuracy of 0.915.
- Good performance on Whole Tumor (WT) and Tumor Core (TC) regions with Dice Score of 0.9136 and 0.9003, respectively.
- Qualitative visualisations indicated tight agreement between estimated and ground truth masks in the majority of scenarios, with slight deviations around small or low-contrast tumor regions

While HybridSegNet++ does not surpass high-capacity 3D transformer-based models on the ET class, it performs better segmentation of TC more than all previously published baseline models and performs competitively on WT. Hence, HybridSegNet++ strikes a good balance between efficiency and accuracy with it excelling in segmenting major tumor areas and being a top player among lightweight 2D models.

6.2 Future Scope

While HybridSegNet++ has good performance, but there are some future directions:

1. **3D Volumetric Context Extension:**

   Adding 3D convolutional layers or 2.5D context windows makes it possible to capture inter-slice information and enhance ET and TC segmentation performance.

2. **Attention Module Integration:**

   Adding self-attention mechanisms in the decoder or across modalities allows the model to better distinguish fine structures and reject irrelevant background.

3. **Dynamic Loss Adjustment:**

   Future deployments can use adaptive loss weighting (e.g., focal loss, Lovasz loss) by epoch-wise class challenge or confidence maps.

4. **Domain Adaptation and Generalization:**

   Model testing against other datasets (e.g., BraTS 2021, private clinical scans) will establish its generalizability and scanner/institution/protocol/patient-group stability.

5. **Post-Processing Improvements:**

   Applying Conditional Random Fields (CRFs) or employing ensemble techniques for post-prediction smoothing would enhance spatial consistency.

REFERENCES

[1] S. Li, X. Sui, X. Luo, X. Xu, Y. Liu, and R. Goh, "Medical Image Segmentation Using Squeeze-and-Expansion Transformers," *arXiv preprint arXiv:2105.09511*, Jun. 2023. https://arxiv.org/abs/2105.09511

[2] W. Yao, J. Bai, W. Liao, Y. Chen, M. Liu, and Y. Xie, "From CNN to Transformer: A Review of Medical Image Segmentation Models," *arXiv preprint arXiv:2308.05305*, 2023. https://arxiv.org/abs/2308.05305

[3] Y. Xie, B. Yang, Q. Guan, J. Zhang, Q. Wu, and Y. Xia, "Attention Mechanisms in Medical Image Segmentation: A Survey," *arXiv preprint arXiv:2305.17937*, 2023.https://arxiv.org/abs/2305.17937

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* (MICCAI), 2015, pp. 234–241.https://arxiv.org/abs/1505.04597

[5] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A Nested U-Net Architecture for Medical Image Segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Springer, 2018, pp. 3–11. https://arxiv.org/abs/1807.10165

[6] D. Jha, M. A. Riegler, D. Johansen, P. Halvorsen, and H. D. Johansen, "ResUNet++: An Advanced Architecture for Medical Image Segmentation," *arXiv preprint arXiv:1911.07067*, 2019.https://arxiv.org/abs/1911.07067

[7] O. Oktay *et al.*, "Attention U-Net: Learning Where to Look for the Pancreas," *arXiv preprint arXiv:1804.03999*, 2018.https://arxiv.org/abs/1804.03999

[8] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, pp. 203–211, 2021.https://www.nature.com/articles/s41592-020-01008-z

[9] Y. Wang, J. Li, Y. Xia, Y. Zhang, and D. Shen, "TransBTS: Multimodal Brain Tumor Segmentation Using Transformer," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* (MICCAI), 2021.https://arxiv.org/abs/2103.04430

[10] H. Cao *et al.*, "Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation," *arXiv preprint arXiv:2105.05537*, 2021. https://arxiv.org/abs/2105.05537

[11] S. Mehta and M. Rastegari, "MobileViT: Light-weight, General-purpose, and Mobile-friendly Vision Transformer," *arXiv preprint arXiv:2110.02178*, 2021.https://arxiv.org/abs/2110.02178

[12] J. Chen *et al.*, "TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation," *arXiv preprint arXiv:2102.04306*, 2021.https://arxiv.org/abs/2102.04306

[13] J. M. J. Valanarasu, P. Oza, I. Hacihaliloglu, and V. M. Patel, "Medical Transformer: Gated Axial-Attention for Medical Image Segmentation," *arXiv preprint arXiv:2102.10662*, 2021.https://arxiv.org/abs/2102.10662

[14] L. Zhu, W. Yang, Y. Chen, B. Zhang, and X. Fan, "Edge-Aware Attention Network for Medical Image Segmentation," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 17, no. 4, pp. 745–756, 2022. https://link.springer.com/article/10.1007/s11548-022-02591-z

[15] M. F. Ahamed, M. R. Islam, M. Nahiduzzaman, M. J. Karim, M. A. Ayari, and A. Khandakar, "A review on brain tumor segmentation based on deep learning methods with federated learning techniques," *Comput. Med. Imaging Graph.*, vol. 103, p. 102313, 2023. https://www.sciencedirect.com/science/article/pii/S0895611123001313

[16] W. Yao, J. Bai, W. Liao, Y. Chen, M. Liu, and Y. Xie, "From CNN to Transformer: A Review of Medical Image Segmentation Models," *arXiv preprint arXiv:2308.05305*, 2023. https://arxiv.org/abs/2308.05305

[17] Y. Xie, B. Yang, Q. Guan, J. Zhang, Q. Wu, and Y. Xia, "Attention Mechanisms in Medical Image Segmentation: A Survey," *arXiv preprint arXiv:2305.17937*, 2023. https://arxiv.org/abs/2305.17937

[18] R. Azad, A. Kazerouni, M. Heidari, E. K. Aghdam, A. Molaei, Y. Jia, A. Jose, R. Roy, and D. Merhof, "Advances in Medical Image Analysis with Vision Transformers: A Comprehensive Review," *arXiv preprint arXiv:2301.03505*, 2023.https://arxiv.org/abs/2301.03505

[19] M. F. Ahamed, M. R. Islam, M. Nahiduzzaman, M. J. Karim, M. A. Ayari, and A. Khandakar, "A review on brain tumor segmentation based on deep learning methods with federated learning techniques," *Comput. Med. Imaging Graph.*, vol. 103, p. 102313, 2023. https://www.sciencedirect.com/science/article/pii/S0895611123001313

[20] L. Zhu, W. Yang, Y. Chen, B. Zhang, and X. Fan, "Edge-Aware Attention Network for Medical Image Segmentation," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 17, no. 4, pp. 745–756, 2022. https://link.springer.com/article/10.1007/s11548-022-02591-z

[21] Y. Wang, J. Li, Y. Xia, Y. Zhang, and D. Shen, "TransBTS: Multimodal Brain Tumor Segmentation Using Transformer," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* (MICCAI), 2021.https://arxiv.org/abs/2103.04430