

**3D BRAIN TUMOR SEGMENTATION USING MULTIMODAL
MRI SEQUENCES AND INCOMPLETE MRI DATA**

A PROJECT REPORT SUBMITTED BY

M.M.D. SAMARANAYAKE

S16533

to the Board of Study in

DEPARTMENT OF STATISTICS AND COMPUTER SCIENCE

in partial fulfillment of the requirement

for the award of the degree of

BSc. (Honors) in Computer Science

of the

UNIVERSITY OF PERADENIYA

SRI LANKA

2021

DECLARATION

I do hereby declare that the work reported in this project report was exclusively carried out by me under the supervision of Dr. Hakim Usoof. It describes the results of my own independent work except where due reference has been made in the text. No part of this project report has been submitted earlier or concurrently for the same or any other degree.

Date: 10.07.2022



.....

Signature of the Candidate

Certified by:

1. Supervisor: Dr. Hakim Usoof

Date: 10.07.2022

Signature:

2. Head of the Department: Dr. Sachith Abeyesundara

Date:

Signature:

REVIEW ON 3D BRAIN TUMOR SEGMENTATION USING MULTIMODAL MRI SEQUENCES AND INCOMPLETE MRI DATA

M.M.D. Samaranayake (S/16/533)

Department of Statistics and Computer Science, University of Peradeniya, Peradeniya, Sri Lanka

Brain tumors has a significant socio-economic burden on mankind and is one of the deadliest types of cancer that can be seen globally. As its dismal prognosis and high cost on treatment makes it abysmal. Therefore, proper identification of brain tumors and treatment planning to reduce the burden on cancer patients is important. If proper in-vivo, non-invasive delineation and identification of glioma structures is in place, treatment design and follow-up can be greatly improved. However, inter-observer variability in tumor segmentation from magnetic resonance images is a major inaccurate source of radiation therapy, necessitating the development of tools to automate this process. This study investigates the optimality of three automated brain tumor segmentation models using 3D Convolutional Neural Networks trained on magnetic resonance images with multiple modality sequences with brain tumors. An ensemble method is also proposed in the study considering a statistical analysis on the models, as to mitigate the segmentation inaccuracies of tumor sub-regions. The study was focused on Cascaded anisotropic, U-Net and Residual-Network convolutional architectures to train the segmentations along with the ensemble model. These models were trained using the Brain Tumor Segmentation Challenge dataset collected for the year 2020. The ensemble model outperformed the rest of the architectures by giving a decent DICE score of 88% of the whole tumor and 75% of the enhancing tumor segmentation. The next part of the study was to identify the missing modality with the highest accuracy drop and reconstruct the particular modality to mitigate the accuracy drop in segmentation. So, the approach was to identify the accuracy by finding the most essential modality combination among the four modalities, which significantly impacts the accuracy of state-of-the-art neural network architectures used. It was learned that FLAIR modality sequence and the T2 modalities had the greatest impact on the segmentation accuracy of the whole-tumor region. The best-found combination was FLAIR, T1 and T2 modalities at it gives the second highest accuracy on segmentation next to all four-modality segmentation accuracy.

ACKNOWLEDGMENTS

First and foremost, I would like to express my deep and sincere gratitude to my supervisor, Dr. Hakin Usoof, for giving me the opportunity to do the research while providing necessary information and giving his invaluable guidance and supervision throughout this research amidst of his tight schedule. His motivation, sincerity and dynamism provided me with great enthusiasm and passion to headway in my research. It was a great honor to work and study under his profound guidance enclosed with great insight. I would like to extend my heartfelt gratitude for his eminent support, and I believe I am greatly indebted to his extraordinary endeavor and support given throughout my research work and in regard.

Also, I would like to thank the staff members of the Department of Statistics and Computer Science for sparing their valuable time and giving suggestions and advice to carry on the research. And also providing the required resources for the fulfillment of the research.

I am extremely thankful to my parents and siblings for their care, sacrifices and prayers and continuing to support me to complete this research.

TABLE OF CONTENTS

DECLARATION	II
ABSTRACT	III
ACKNOWLEDGMENTS	IV
TABLE OF CONTENTS	V
LIST OF FIGURES	VII
LIST OF TABLES	VIII
LIST OF ABBREVIATIONS	IX
CHAPTER 01: INTRODUCTION	1
1.1 Problem Statement	2
1.2 Objectives	3
CHAPTER 02: BACKGROUND	4
2.1 Glioma Detection	4
2.2 MRI Image Acquisition	5
2.3 MRI Modality Image Acquisition	6
CHAPTER 03: LITERATURE REVIEW	9
3.1 CNNs for Semantic Segmentation	10
3.2 Brain Tumor Segmentation	11
CHAPTER 04: METHODOLOGY	13
4.1 Dataset	13
4.2 Data Preprocessing	14
4.3 Pipeline for Data Ingestion	14
4.4 Network Architecture	18
4.4.1 CA-CNN (Cascaded Anisotropic Convolutional Neural Network)	18
4.4.1.1 Cascading Framework	18
4.4.1.2 Anisotropic Convolutional Neural Networks	19
4.4.1.3 Anisotropic and Dilated Convolution Layer	20
4.4.1.4 Residual Connection Layer	21
4.4.1.5 Multi-scale Predictions	21
4.4.1.6 Multi View Fusion	22
4.4.2 3D U-Net	22
4.4.2.1 Dense Blocks	23
4.4.2.2 Residual-Inception Blocks	24
4.4.3 Pre-Trained Res-Net	25
4.4.3.1 Residual Connection	26
4.4.3.2 Encoder	26
4.4.3.3 Decoder	27
4.4.4 Ensemble Model	28
4.5 Implementation	29
4.6 Experimental Design	30
4.6.1 Dataset Splits	30
4.6.2 Evaluation	30
4.6.3 Activation Function and Loss Function	31
4.6.4 Training Details	33
4.6.5 Modality Combinations	33

4.7 Testing Environment	34
4.7.1 Software specification	34
4.7.2 Hosted Hardware specifications	34
4.7.3 Programming Technologies	34
CHAPTER 05: RESULTS AND DISCUSSION	35
5.1 Overview of the Results and Discussion	35
5.2 Segmentation Results on all the CNN Architectures	36
5.3 Modality Combination Experimentation on the Ensemble	38
5.3.1 Segmentation Results on All four modalities	38
5.3.2 Segmentation Results of FLAIR, T1, and T1-ce Modalities with the Ensemble	40
5.3.3 Segmentation Results of FLAIR, T1, and T2 Modalities with the Ensemble	42
5.3.4 Segmentation Results of FLAIR, T2, and T1-ce Modalities with the Ensemble	44
5.3.5 Segmentation Results of T1, T2, and T1-ce Modalities with the Ensemble	46
5.4 Predicted Segmentations vs Ground Truth	48
CHAPTER 06: CONCLUSIONS	50
CHAPTER 07: FUTURE WORK	51
7.1 More effective loss function for training deep CNNs	51
7.2 Survival Prediction Feature Extraction	51
7.3 Generative Adversarial Network for missing modality Reconstruction	51
CHAPTER 08: REFERENCES	52
CHAPTER 09: APPENDIX	57

LIST OF FIGURES

Figure 2. 1: Brain MRI Multimodal Scans Representation	8
Figure 4. 1: Transformation of Image data in the data ingestion pipeline	16
Figure 4. 2: The proposed triple cascaded framework for brain tumor segmentation.	19
Figure 4. 3: Cascaded Anisotropic Convolutional Networks Architecture	20
Figure 4. 4: Illustration of multi-view fusion at one level of the proposed cascade.	21
Figure 4. 5: Proposed 3D U-Net Architecture	23
Figure 4. 6: Densely Connected Structure of the first block of encoder path	24
Figure 4. 7: Proposed RIB structure of the 3D UNet	25
Figure 4. 8: Residual learning: a building block	25
Figure 4. 9: Pre-Trained Res-Net structure.	26
Figure 4. 10: Structure of each block unit in Fig.4.9	27
Figure 4. 11: Axial representation of an example patch of size 80 x 80 x 80.	28
Figure 4. 12: Proposed Ensemble Model for Tumor Segmentation	29
Figure 4. 13: Tumoral structures of the 3D brain tumor.	31
Figure 4. 14: Jaccard Index depicted with an	32
Figure 5. 1: Results Workflow	35
Figure 5. 2: Evaluation graphs on all the modalities	38
Figure 5. 3: Dice and Jaccard from Validation for all modalities	39
Figure 5. 4: Evaluation graphs on FLAIR, T1, and T1-ce	40
Figure 5. 5: Dice and Jaccard from Validation for FLAIR, T1, and T1-ce	41
Figure 5. 6: Evaluation graphs on FLAIR, T1, and T2	42
Figure 5. 7: Dice and Jaccard from Validation for FLAIR, T1, and T2	43
Figure 5. 8: Evaluation graphs on FLAIR, T2, and T1-ce	44
Figure 5. 9: Dice and Jaccard from Validation for FLAIR, T2, and T1-ce	45
Figure 5. 10: Evaluation graphs on T2, T1, and T1-ce	46
Figure 5. 11: Dice and Jaccard from Validation for T2, T1, and T1-ce	47
Figure 5. 12: Predicted output of a Patient's Brain tumor segmentation	48
Figure 5. 13: Ground truth of a Patient's Brain tumor	48
Figure 5. 14: 3D ground truth of a Brain tumor	49
Figure 5. 15: 3D prediction of a Brain tumor	49
Figure 9. 1: Residual Network Architecture	57
Figure 9. 2: U-Net Architecture	58
Figure 9. 3: Cascaded Anisotropic Architecture	59
Figure 9. 4: Ensemble Architecture	60

LIST OF TABLES

Table 2. 1: Image Acquisition Time Requirements	6
Table 4. 1: Dataset Contents	13
Table 5. 1: Segmentation results on every model architecture	36
Table 5. 2: Segmentation Results on all 4 modalities	39
Table 5. 3: Segmentation Results on FLAIR, T1, and T1-ce Combination	40
Table 5. 4: Segmentation Results on FLAIR, T1, and T1-ce	42
Table 5. 5: Segmentation Results on FLAIR, T2, and T1-ce	44
Table 5. 6: Segmentation Results on T1, T2, and T1-ce	46

LIST OF ABBREVIATIONS

GTR:	Gross Total Resection
STR:	Sub Total Resection
MRI:	Magnetic Resonance Imaging
GPU:	Graphics Processing Unit
TPU:	Tensor Processing Unit
ETL:	Extract, Transform & Load
CNN:	Convolutional Neural Network
FCN:	Fully Connected Network
RIB:	Residual-Inception Block
DNA:	Deoxyribonucleic Acid
CSF:	Cerebrospinal Fluid
CA-CNN:	Cascaded Anisotropic Convolutional Neural Network
PT:	Pre-Trained
API:	Application Programming Interface

CHAPTER 01

INTRODUCTION

There are millions upon millions of cells in the brain. These cells are generated from the time of conception until roughly the age of seven. These brain cells are never meant to divide again after they stop dividing. As you can see, brain cell division is strictly regulated and controlled. When this control is lost in a single cell, the cell begins to undergo cell division uncontrollably. All evidence from cancer research indicates that the disease is caused by a rapidly dividing cell that lacks regulators to prevent it from dividing. Normal human cells grow, and function based mainly on the information in each cell's DNA. Brain and spinal cord tumors, like other tumors, are caused by changes in the DNA inside cells. DNA is the chemical that makes up our genes, which control how our cells function. We usually look like our parents because they are the source of our DNA. However, DNA affects more than how we look. Some genes control when our cells grow, divide into new cells and die. Specific genes that help cells grow, divide, and stay alive are called oncogenes. Genes that help keep cell division under control, repair mistakes in DNA, or make cells die at the right time are called tumor suppressor genes. Cancers can be caused by DNA changes that turn on oncogenes or turn off tumor suppressor genes [1]. These gene changes can be inherited from a parent, but more often, they happen during a person's lifetime[2].

Brain tumors can be either malignant or benign, and both are harmful to the brain. A malignant brain tumor can penetrate and destroy vital tissues and cells. These tumors have the potential to spread to other regions of the body. On the other hand, benign tumors are slow-growing cells that rarely spread to other sections of the body. When it comes to the brain, both of these tumors are deadly. Because the skull cannot expand while the tumor grows, pressure is applied to the brain. The brain tissue gets affected as a result of the strain. Both types of brain cancer can be fatal if left untreated.

According to the IARC, the number of new cases reported in 2020 for both genders and every age group was 308,102. The number of deaths reported was 251 329 out of the World population. New patients are diagnosed with brain tumors every year. The average survival rate for a patient in five years is one in three. Early identification of cancer is critical to prevent it from progressing to a life-threatening stage. The strongest predictor of long-term survival is detecting a tumor while it is still in its early stages.

Doctors utilize a variety of tests to locate and diagnose a brain tumor and determine the type of tumor. They also perform tests to see if cancer has migrated to other parts of the body from where it began. This is known as metastasis, and it is uncommon in the case of a primary brain tumor. In most cases, MRI is used to diagnose a brain tumor. Brain tumor segmentation entails diagnosing, delineating, and distinguishing tumor tissues such as active cells, necrotic cores, and edema from normal brain tissues such as Gray Matter, White Matter, and CSF. MRIs can help with diagnosis, tumor classification and treatment planning by revealing some main indications of unhealthy tissue[3].

Brain tumor segmentation necessitates expert knowledge in order to identify unhealthy from healthy tissues, those tasks are both costly and time-consuming. The use of computer-assisted automatic segmentation methods can help with diagnosis. Current high segmentation performances obtained by deep learning methods make them good candidates for achieving this task. In this research, the main focus would be on detecting brain tumors with incomplete MRI data using deep learning compared to state-of-the-art[4] model approach with complete MRI data.

1.1 Problem Statement

The research community has focused its efforts on the brain tumor segmentation problem, and initiatives such as the Brain Tumor Segmentation (BraTS) challenge[5] have made the problem more accessible to a wider audience. As a result, a wide range of computer algorithms for automating the delineation of brain tumors have been developed.

In the field of brain tumor segmentation using deep learning techniques, it is evident that there are a numerous models and techniques for accurate automatic segmentation of brain tumors. In the segmentation process it is required to have multi modalities of MR images for the detection of the tumor with higher accuracy in 3D segmentation. These modalities play a major role in identifying the minute details of the tumor during the segmentation process.

1.2 Objectives

The overall objective of the research is as follows.

- Using MRI modalities T1, T2, T1-Gd, and FLAIR train a CNN with different modality combinations for tumor segmentation and finding the most essential modality combination, which significantly impacts the accuracy of state-of-the-art NN architectures used for brain tumor image segmentation and exploring the techniques needed to reconstruct the missing MRI data to mitigate the accuracy drop compared to state-of-the-art methods.

CHAPTER 02

BACKGROUND

Although brain tumors are rare, they are one of the deadliest cancers diagnosed among people worldwide. Primary and metastatic brain tumors are two brain tumors that must be distinguished while diagnosing a brain tumor. Primary tumors begin in brain tissue cells, whereas metastatic cancers begin in any body region and move to the brain. Gliomas are tumors that originate from glial cells in the brain. They are the most common brain tumors that focus on current brain tumor identification research. Gliomas are classified as low-grade gliomas such as Astrocytoma and Oligodendrogliomas to high-grade (grade IV) Glioblastoma Multiforme, which is the most severe and common primary malignant brain tumor[3].

2.1 Glioma Detection

Glioma detection at an early stage is critical for better treatment options. Computed Tomography (CT), Single-Photon Emission Computed Tomography (SPECT), Position Emission Tomography (PET), Magnetic Resonance Spectroscopy (MSR), and Magnetic Resonance Imaging (MRI) are all medical imaging techniques that are used to provide valuable information about the shape, size, location, and metabolism of brain tumors. While these features are used in combination to provide the highest detailed information about brain tumors, due to its soft-tissue contrast and wide availability, MRI is considered the standard technique[6].

Doctors can use an MRI to check your organs, tissues, and skeletal system. It creates high-resolution images of the inside of the body that can be used to identify a wide range of issues. It is a non-invasive Vivo imaging technique that employs radio frequency signals to excite target tissue and generate internal images under the influence of a potent magnetic field. MRI sequence images are created by varying excitation and repetition during image acquisition. These various MRI modalities generate different types of tissue contrast images, which provide valuable structural information and allow for the diagnosis and segmentation of tumors and their sub-regions.

2.2 MRI Image Acquisition

The magnetic characteristics of atomic nuclei are used in MRI. Protons that are ordinarily randomly orientated within the water nuclei of the tissue being investigated are aligned using a high, uniform external magnetic field. After that, an external Radio Frequency (RF) energy is used to disrupt or break the alignment (or magnetization). RF energy is emitted as the nuclei return to their resting alignment through various relaxation processes. The emitted signals are measured after a specific amount of time has passed from the first RF. The Fourier transformation is used to convert the signal's frequency information from each place in the scanned plane to matching intensity levels, which are then represented as shades of gray in a matrix of pixels. Different types of images can be formed by altering the sequence of RF pulses applied and collected. The duration between subsequent pulse sequences delivered to the same slice is known as the **Repetition Time** (TR). The period between the delivery of the RF pulse and the reception of the echo signal is known as the **Time to Echo** (TE) [7].

Tissue can be characterized by two different relaxation times T1 and T2.

- T1 (longitudinal relaxation time) is the time constant that determines the rate at which excited protons return to equilibrium. It is a measure of the time taken for spinning protons to realign with the external magnetic field.
- T2 (transverse relaxation time) is the time constant that determines the rate at which excited protons reach equilibrium or go out of phase with each other. It measures the time taken for spinning protons to lose phase coherence among the nuclei spinning perpendicular to the main field.

Standard MRI modalities for GLIOMA diagnosis:

There are different modalities of MR images, each of these MRI modalities capture different characteristics of the underlying anatomy of the brain. Some of these modalities are used in referenced for clinical diagnosis of brain tumors. Mainly T1 (spin lattice relaxation)

2.3 MRI Modality Image Acquisition

- T1-weighted MRI (T1)
 - T1-weighted images are produced by using short TE and TR times. The T1 properties of tissue predominately determine the contrast and brightness of the image.
- T2-weighted MRI (T2)
 - T2-weighted images are produced by using longer TE and TR times. In these images, the contrast and brightness are predominately determined by the T2 properties of tissue.

In general, T1- and T2-weighted images can be easily differentiated by looking at the CSF. CSF is dark on T1-weighted imaging and bright on T2-weighted imaging.

- T1-weighted MRI with Gadolinium Contrast Enhancement (T1-Gd)
 - T1-weighted imaging can also be performed while infusing Gadolinium (Gad). Gad is a non-toxic paramagnetic contrast enhancement agent. When injected during the scan, Gad changes signal intensities by shortening T1. Thus, Gad is very bright on T1-weighted images. Gad enhanced images are instrumental in looking at vascular structures and breakdown in the blood-brain barrier [e.g., tumors, abscesses, inflammation (herpes simplex encephalitis, multiple sclerosis, etc.)].
- Fluid Attenuated Inversion Recovery (FLAIR)
 - The Flair sequence is similar to a T2-weighted image except that the TE and TR times are very long. By doing so, abnormalities remain bright, but normal CSF fluid is attenuated and made dark. This sequence is very sensitive to pathology and makes the differentiation between CSF and an abnormality much easier.

Table 2. 1: Image Acquisition Time Requirements

	TR (msec)	TE (msec)
T1-Weighted (short TR and TE)	500	14
T2-Weighted (long TR and TE)	4000	90
Flair (very long TR and TE)	9000	114

In recent years, many publications applying computer vision techniques to static medical imagery have grown from hundreds to thousands [8]. Because of the visual pattern-recognition nature of diagnostic tasks in these specialties and the growing availability of highly structured images, a few areas have received significant attention: radiology, pathology, ophthalmology, and dermatology. However, the distinct features of medical imagery present several challenges to DL-based computer vision. Images can be enormous. For example, digitizing histopathology slides yields Gigapixel images with approximately 100000x100000 pixels, whereas typical CNN image inputs have approximately 200x200 pixels.

Furthermore, different chemical preparations will result in different slides for the same piece of tissue, and different digitization devices or settings will result in different images for the same slide. MRI produces equally massive 3D images, forcing standard CNNs to either work with a set of 2D slices or modify their internal structure to process in 3D. Similarly, ultrasound produces a time series of noisy 2D slices from a 3D context—slices spatially correlated but not aligned. DL has begun to take into account the unique challenges of medical data [9].

Although the number of slices of 2D images produced during MRI acquisition varies depending on the device, approximately 150 slices of 2D images are produced to represent the 3D brain volume. Furthermore, when the required standard modalities' slices are combined for diagnosis, the data becomes highly populated and complicated [3].

T1 images are typically used to differentiate healthy tissues, whereas T2 images are used to delineate the edema region, producing a bright image signal. The bright signal of the accumulated contrast agent (gadolinium ions) in the active cell region of the tumor tissue in T1-Gd images easily distinguishes the tumor border. Because necrotic cells do not interact with the contrast agent, they can be seen in the hypo-intense part of the tumor core, allowing them to be easily distinguished from the active cell region on the same sequence. In addition, the signal of water molecules is suppressed in FLAIR images, which aids in distinguishing the edema region from the CSF [8].

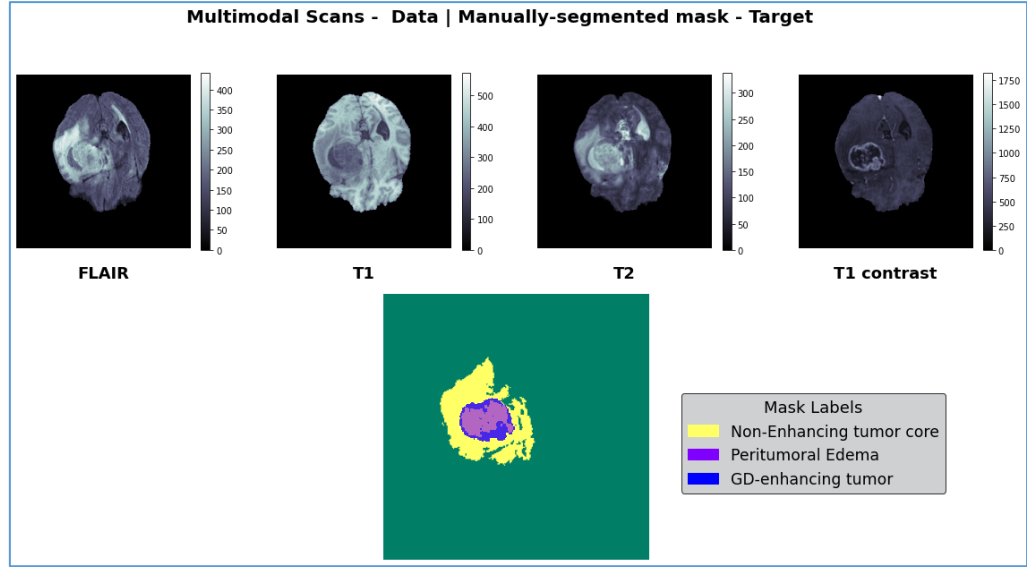


Figure 2. 1: Brain MRI Multimodal Scans Representation

Multimodal scans of a brain MRI are depicted in the Figure 2.1 which are the modalities required for segmentation of the tumor.

Multi-layer neural network-based automatic detection approaches have been extensively used to segment brain tumors. The MICCAI Brain Tumor Segmentation (BraTS) challenge gathered the most publicly available brain tumor segmentation dataset to date. They employ well-defined training and testing splits, allowing us to compare different approaches fairly. Currently, state-of-the-art results on the BraTS dataset can produce results that are comparable to those of human experts [10].

However, as provided in the dataset, the current brain tumor segmentation benchmark requires complete MR images in all T1, T2, T1c, and Flair modalities. To differentiate tissues anatomies, and multi-modality images provide specific data. Brain tumor segmentation from MRIs requires learning a complex function that works across all four modalities and is impacted by anatomical morphology and local pixel values. However, different hospitals may use different procedures and practices when performing MRIs in reality. It is normal to be missing one of the four modalities. This may present some difficulties when four complete modalities are required on MR images at the same time.

CHAPTER 03

LITERATURE REVIEW

Manual segmentation of brain tumors from large amounts of MRI images generated in the clinical routine is time-consuming and challenging. Automatic segmentation using deep learning methods has recently gained popularity because these methods achieve state-of-the-art results and can address this problem better than other methods. Deep learning methods can also be used to process and evaluate large amounts of MRI-based image data efficiently and objectively [11].

Glioblastoma plays an essential role in planning correlated surgeries and treatments for brain tumors. The conventional models of survival prediction rely on radiomic features using MRI. Some proposed approaches were evaluated by comparative experiments with state-of-the-art models in synthesis, segmentation, and overall survival (OS) prediction [11]. Observations have been made that adding missing MRI modality improves the segmentation prediction, and expression levels of gene markers have a high contribution in the GBM prognosis prediction, and fused radio genomic features boost the OS estimation.

Automation of brain tumor segmentation in 3D MRIs is vital to assess the diagnostic and treatment of the disease. High memory consumption is a problem in 3D-CNNs. Moreover, most methods do not include uncertain information, which is especially critical in medical diagnosis. One of the current research studies 3D encoder-decoder architectures trained with patch-based techniques to reduce memory consumption and decrease unbalanced data [12]. They introduce voxel-wise uncertainty information, both epistemic and aleatoric using test-time dropout (TTD) and data-augmentation (TTA), respectively. In addition, a hybrid approach that helps increase the accuracy of the segmentation was also developed.

To improve the performance and reduce the complexity involves in the medical image segmentation process, Berkeley wavelet transformation (BWT) based brain tumor segmentation has been used[13]. Furthermore, to improve the accuracy and quality rate of the SVM-based classifier, relevant features are extracted from each segmented tissue. The experimental results of the proposed technique have been evaluated and validated for performance and quality analysis on magnetic resonance brain images, based on accuracy, sensitivity, specificity, and dice similarity index coefficient.

The experimental results have achieved 96.51% accuracy, 94.2% specificity, and 97.72% sensitivity, demonstrating the effectiveness of the proposed technique for identifying normal and abnormal tissues from brain MR images. The experimental results have obtained an average of 0.82 dice similarity index coefficient, indicating better overlap between the automated (machines) extracted tumor region with the manually extracted tumor region by radiologists.

A study has proposed a computer-aided detection approach to diagnose brain tumors in their early stage using Mathematical Morphological Reconstruction (MMR). Image is pre-processed to remove noise and artifacts and then segmented to find regions of interest with probable tumors. Many textural and statistical features are extracted from the segmented image to classify whether the brain tumor in the image is benign or malignant. Experimental results show that the segmented images have a high accuracy while substantially reducing the computation time. The study shows that the proposed solution can be used to diagnose brain tumors in patients with a high success rate[14].

3.1 CNNs for Semantic Segmentation

Some of the early works in semantic segmentation using CNNs make use of their representation learning capabilities by learning to predict the class of each pixel centered around a patch extracted from the raw image [15], or by using a multiscale convnet trained to extract dense feature vectors to encode regions of multiple sizes centered around each pixel [16]. However, the work by [17] laid the groundwork for semantic segmentation with CNNs, on which most, if not all, state-of-the-art methods are based. In this paper, they present the first case of end-to-end dense training with whole images using fully convolutional versions of existing networks like [18] and [19].

Backwards convolution (also known as transpose convolution or deconvolution) is introduced, as is a layer that learns an interpolation filter to up sample coarse outputs to dense pixels, and skip-connections, which connect lower layers with higher layers to combine local predictions (lower layers) with global structure (higher layers). Skip-connections are an important component of [20], which describes an architecture with a contracting path that captures context, a symmetric expanding path that enables localization, and skip-connections between the symmetric elements in the contracting and expanding paths that allow for a combination of features from both.

This model won the ISBI challenge for neuronal structure segmentation in electron microscopic stacks, and it has since become a reference model for medical segmentation tasks [21] and [22].

Some studies have raised the issue of repurposing CNNs designed for image classification to perform dense prediction in semantic segmentation. In [23], they propose a novel architecture for dense prediction tasks that involves passing the input image through a stack of dilated convolutions with increasing dilation rate. Dilated convolutions support exponential expansion of the receptive field without loss of coverage, this approach allows for the aggregation of multiscale contextual information without loss of resolution. Alternatively, in [24], they design an architecture with two processing streams: one that processes images at full resolution and enables precise boundary delineation, similar to the previously presented work by [23], and another that presents the standard encoder-decoder architecture that extracts image representations for robust recognition. They associate the extracted features from each processing stream with Full Resolution Residual Units, which combine them in a residual approach.

3.2 Brain Tumor Segmentation

Several supervised learning methods have previously been used to tackle the task of brain tumor segmentation. One of the early lines of research for this topic was the use of voxel-wise classifiers, such as Random Forests, on hand-crafted features. The combination of texture-based features with an adapted Markov Random Field (MRF) implementation [25] resulted in one of the best performing models in the BraTS Challenge 2013 [26]. The proposed algorithm is divided into two stages: in the first, texture features obtained from multi-spectral Gabor filters are used to coarsely segment the tumor; in the second, the coarse boundaries are refined at the voxel level using a modified MRF framework that carefully separates different tumor subclasses from each other and from healthy tissue. Another successful approach in the BraTS Challenge 2013 was proposed by [26]. They also propose a two-stage procedure in which they first obtain a set of supervoxels by clustering voxels based on similarity and proximity, and then refine this over-segmentation with an MRF. Although these methods have been very successful, some argue that their modeling capabilities are still severely limited.

As a result, and as a result of the emergence of deep learning techniques in artificial vision [27], the medical image analysis community has gradually transitioned from systems that use hand-crafted features to systems that learn features from data. These methods offer a powerful alternative for supervised learning, with high model capacity and the ability to learn highly discriminative features, and have achieved near-human performance in some cases [28]. CNNs underpin the vast majority of deep learning models used in brain tumor segmentation. In [29], they look into the possibility of using CNNs to directly segment tumors in brain images.

They apply commercially available 2D CNNs to axial slices of the brain, claiming that by performing 2D convolutions rather than 3D convolutions, the method is more computationally efficient. The 2D-CNN approach is used in subsequent works, such as [30]. The main contribution of this work is threefold: first, they propose a novel two-pathway architecture that learns about the local details of the brain (local pathway) as well as the larger context (global pathway); second, they propose a two-stage training procedure that they believe is critical to dealing with the problem's unbalanced nature; and third, they propose a cascade architecture with different configurations that replaces more complex structured output methods, such as (CRF).

CHAPTER 04

METHODOLOGY

The main requirement behind the research was to identify the highest impacting modality on the tumor segmentation accuracy. So, several models were trained using different modality combinations to extract the high-level features from the multi modalities and identify the highly impactful modality on the segmentation accuracy. The dataset and preprocessing steps will be described in this section, and the research methodology will be discussed.

4.1 Dataset

In this study of brain tumor segmentation on different MRI modalities, the BraTS Challenge 2020 dataset was utilized to evaluate the models' performance [5]. The dataset contains a training set, a validation set of MRI images, and numerical and categorical data. In addition, this dataset contains the patients' resection status and the patients' survival days in the training dataset.

Table 4. 1: Dataset Contents

	Training Data			Validation Data		
MRI images	369 Patients			125 Patients		
Survival Information	237 Patients	GTR	119	29 Patients	GTR	29
		STR	10		STR	0
		Unidentified	107			

The dataset contains 369 multimodal clinically acquired pre-operative MRI scans. Each multimodal is formed by T1, T1-Gd, T2, and T2-FLAIR volumes acquired with various scanners from 20 institutions worldwide. All scans were manually segmented by one to four raters and approved by experienced neuroradiologists. The ground-truth labels include the enhancing tumor, peritumoral edema, necrotic and non-enhancing tumor with visible tumors.

4.2 Data Preprocessing

In the BraTS challenge dataset, each multi-modal scan is co-registered to the same anatomical template, skull-stripped, and resampled to 1mm3 isotropic resolution. As a result, no additional preprocessing is required, and the dataset can be used as-is for training and evaluation. To simplify the dataset, the labeling was changed: assigned Edema to label 1, Enhancing tumor to label 2 and Necrosis and Non-enhancing tumor to label 3.

To assure the quality of the dataset, procedures were taken into account before preprocessing was done. This procedure includes extracting the axial plane images for each modality and visually inspecting for defects or misaligned features. The data quality defects found in the dataset were:

- **Empty images and missing annotations.**
- **Difference in resolution** between acquisitions of the same subject.
- **MR imaging modalities were not aligned** with respect to the others, i.e., they were not in the same subject space. These cases were discarded because a CNN requires aligned grid-like data to learn the relationship between tissue appearances in tumoral structures.

However, the intensities of all the images were normalized to reduce the bias. The intensity value of each MRI is subtracted from the mean and divided by the brain region's standard deviation. To reduce overfitting in the model, random flipping and Gaussian noise were used for data augmentation in the training dataset. This helps the model to generalize well by removing unwanted biasing. The normalized intensity value I_n of a slice is computed in equation (1) by using the normalization equation, where I is the original intensity value of the slice, μ is the intensity value of I , and σ is the standard deviation of I , given as:

$$I_n = \frac{I - \mu}{\sigma} \quad (1)$$

4.3 Pipeline for Data Ingestion

The set of operations that aim to load data from a source, transform it to an appropriate format for the learning algorithm (in this case, a CNN), and then feed it to such algorithms to perform computations. It is regarded as a standard Extract, Transform, Load (ETL) process that ensures data quality, consistency, and conformity to the expected data format.

Google Colab resources were used to retrieve image data stored in google drive as it contained over 30GB of images and was used for model building and training. As Google Colab had GPU and TPUs for accelerated processing, it took less time to run several epochs on the model [31].

The pipeline begins by listing the available data sources, in this case, the subjects for the current phase of the execution process (train, evaluation and inference). These data sources are then added to a queue, which can be shuffled if necessary. A mapping function is applied to these sources to represent the Transform operations. This mapping function is in charge of reading the data from a CSV file, finding the path to the image's location, and optionally performing preprocessing and data augmentation operations and outputting the data in the format expected by the image processing algorithm. To optimize the parallelism of the ETL process, the transformed data is optionally reshuffled, and then a prefetching operation is performed.

Finally, the data is organized into batches and loaded into Google Colab resource, typically an accelerated computing device like a GPU or a TPU. The entire ETL process is designed to be asynchronous from the start. Each of these operations (extract, transform, and load) can be executed in parallel, avoiding the computational bottlenecks that would occur in a sequential execution scenario. This is especially true in accelerated computing environments, where data availability for the computing device is critical for optimal training and inference.

The MR images are read from google drive, along with their corresponding ground truths, and then undergo a series of transformations, categorized as preprocessing and augmentation, to produce a data structure that will be fed to the learning algorithm later in the process.

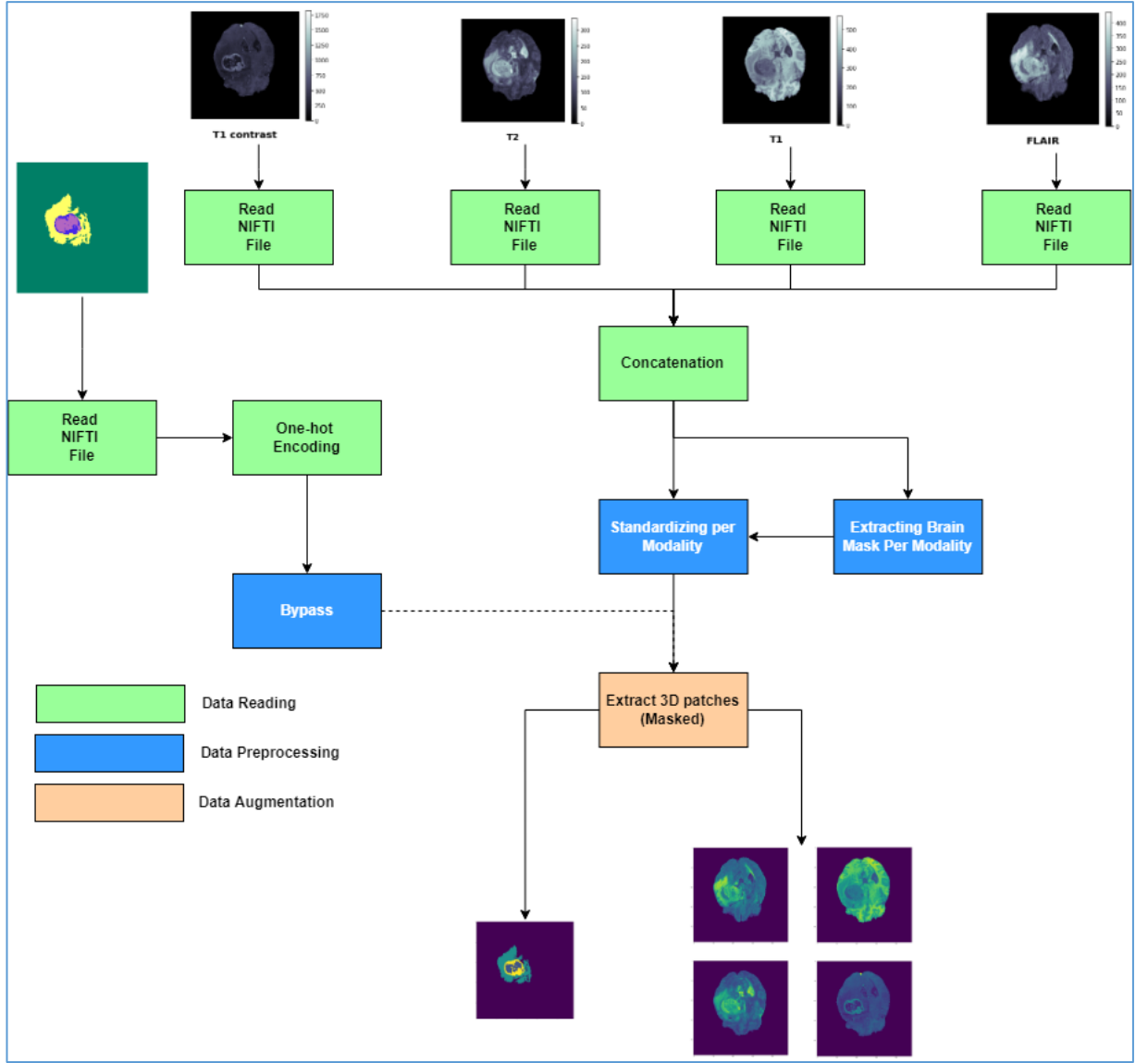


Figure 4. 1: Transformation of Image data in the data ingestion pipeline

The transformation operations depicted in Fig 4.1 require special attention because they are precise to this task. All the MRI modalities are read from the Google drive along with the ground truth containing the segmentation in NIFTI(.nii) format. These modalities are in a 3D tensor data array. These images are concatenated along the last axis, which in turn creates a 4D tensor having dimensions $W \times H \times D \times M$, where W is the width, H is the height, D is the depth and M is the number of MR modalities used in concatenation. The data appears to be 2D but is a 3D tensor with dimensions $W \times H \times 3$, where the three channels represent the color channels in the RGB image.

The labels, on the other hand, are one-hot encoded: from a 3D tensor in which each voxel contains a single numeric value representing its voxel label (one of 0, 1, 2, or 3), a 4D tensor of dimensions $W \times H \times D \times 4$ is created with a 1 in the axis of the last dimension that corresponds to that voxel's label value. For example, if we take a voxel located at a position x, y, z , which gives the label 2, after using one-hot encoding, it will be represented as $l_{x,y,z} = [0,0,1,0]$.

A brain mask is extracted after image concatenation using simple thresholding: any value greater than 0 is considered brain. This is effective because all the data are skull-stripped, masking all values outside the brain. We can perform a standardization operation over each modality individually, only considering the voxels that belong to the brain, thanks to this extracted mask and the fact that the data is aligned.

Finally, a 3D patch extraction methodology was employed to generate brain data patches for training and evaluation of the model. There are two reasons for such data augmentation mechanisms. First, we reduce computational and memory requirements by using smaller patches rather than whole brain volumes.

As a result, larger models can be trained with less time, and batch sizes larger than one can be used in current-generation GPUs. Second, patch extraction provides a mechanism for balancing the class representation in this problem, which is highly unbalanced. No matter how large a tumor is, it will always represent a small percentage of the brain's total volume. This is especially true for intra-tumoral structures like enhancing tumors.

The patch extraction algorithm takes as input the volume of the entire brain and, optionally, the volume of the ground truth annotations (also representing the entire brain) and outputs a set of patches whose centers are chosen at random and, in the case of providing the ground truth annotations, which are based on the desired distribution of class representation. The patch size and class representation distribution are parametrized, which means we can extract smaller or larger patches and over-represent or under-represent specific classes. It also accepts a mask that specifies the set of valid locations from which the patches' centers can be chosen randomly. It should be noted that, while a patch can be centered in a voxel containing one of the intra-tumoral structures, the remaining voxels of the patch do not necessarily belong to tumoral tissues.

The opposite is also true: while the center voxel of a patch may belong to the background, the other voxels may belong to tumoral structures (for example, when the central voxel is located near the tumor's boundary). This modifies the class representation to reflect the true distribution of class probabilities, despite the fact that we artificially change it by under and over sampling specific classes, which provides a balance between specificity and sensitivity.

4.4 Network Architecture

All of the Convolutional Architectures implemented and trained in this work are 3D and fully convolutional by design, which means they can be trained using 3D patches of data, and then used for inference with whole-brain volumes. In addition, an ensemble model comprised of three different convolutional neural network architectures was used to perform accurate and robust brain tumor segmentation.

For tumor segmentation, several models have been proposed. They differ in model depth, the filter number, connection method, and other factors. As a result, model performance and behavior can differ depending on the model architecture. Model variance can be reduced, and overall performance improved by training different models separately and merging the results. In this study, three different CNN models were fused, and the ensemble was created using a majority (voting) rule. The following sections provide a detailed description of each model.

4.4.1 CA-CNN (Cascaded Anisotropic Convolutional Neural Network)

4.4.1.1 Cascading Framework

Fig.4.2 depicts the proposed cascaded framework. We use three networks to hierarchically and sequentially segment brain tumor substructures, each of which deals with a binary segmentation problem. The first network (WNet) is responsible for segmenting the entire tumor from multi-modal 3D volumes of the same patient. The tumor's bounding box is then obtained.

The cropped region with the bounding box of the input images was used as the input of the second network (TNet) to segment the tumor core. Similarly, the image region within the tumor core's bounding box was used as the input of the third network (ENet) to segment the enhancing tumor core. The bounding boxes were generated automatically during the training stage based on the ground truth. The bounding boxes were generated in the testing stage based on the binary segmentation results of the entire tumor and the tumor core, respectively [4].

WNet's segmentation result was used as a crisp binary mask for TNet's output, and TNet's segmentation result was used as a crisp binary mask for ENet's output, which serves as anatomical constraints for the segmentation [4].

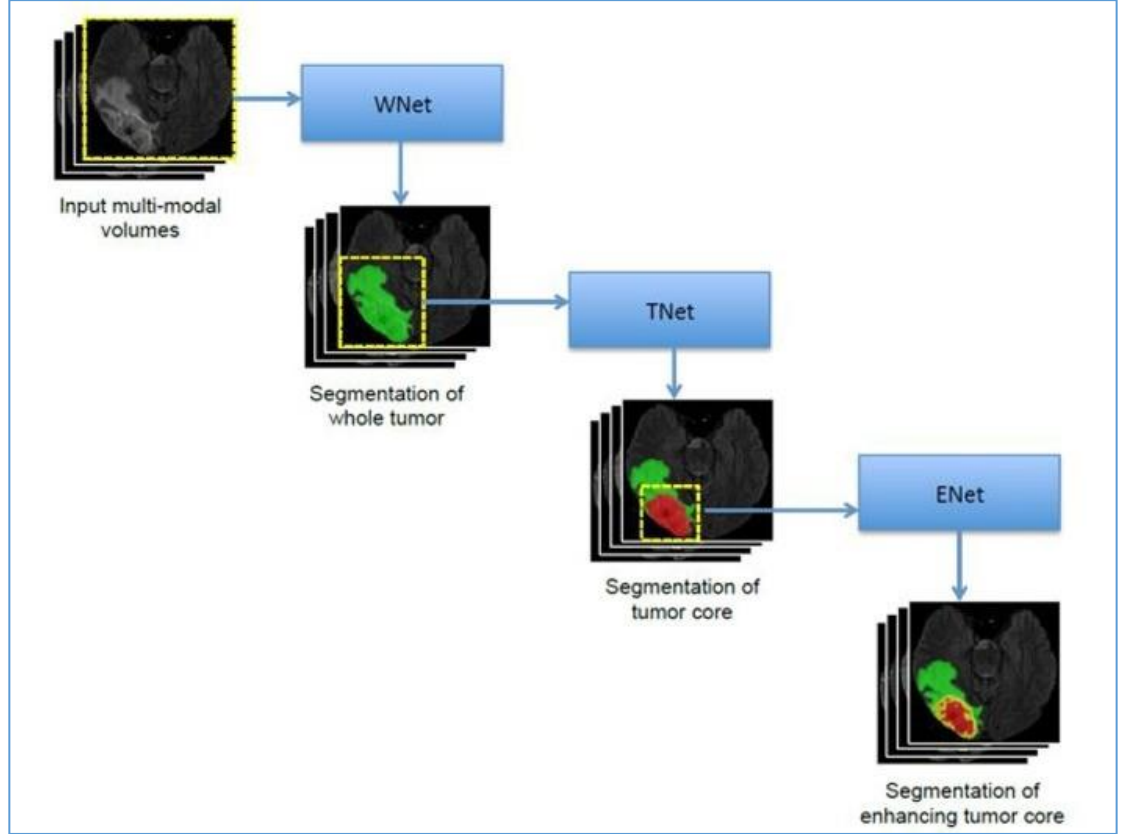


Figure 4. 2: The proposed triple cascaded framework for brain tumor segmentation.

4.4.1.2 Anisotropic Convolutional Neural Networks

The balance of the receptive field, model complexity, and memory consumption should be considered for 3D neural networks. A small receptive field forces the model to use only local features, whereas a larger receptive field allows the model to learn more global features. Many 2D networks, such as FCN and U-Net, use a large receptive field to capture features from the entire image context. For training and testing, they require a large patch size. Using a large 3D receptive field also aids in the acquisition of more global features for 3D volumes. The resulting large 3D patches for training, however, consume much memory, limiting the resolution and number of features in the network, resulting in limited model complexity and lower representation ability [32].

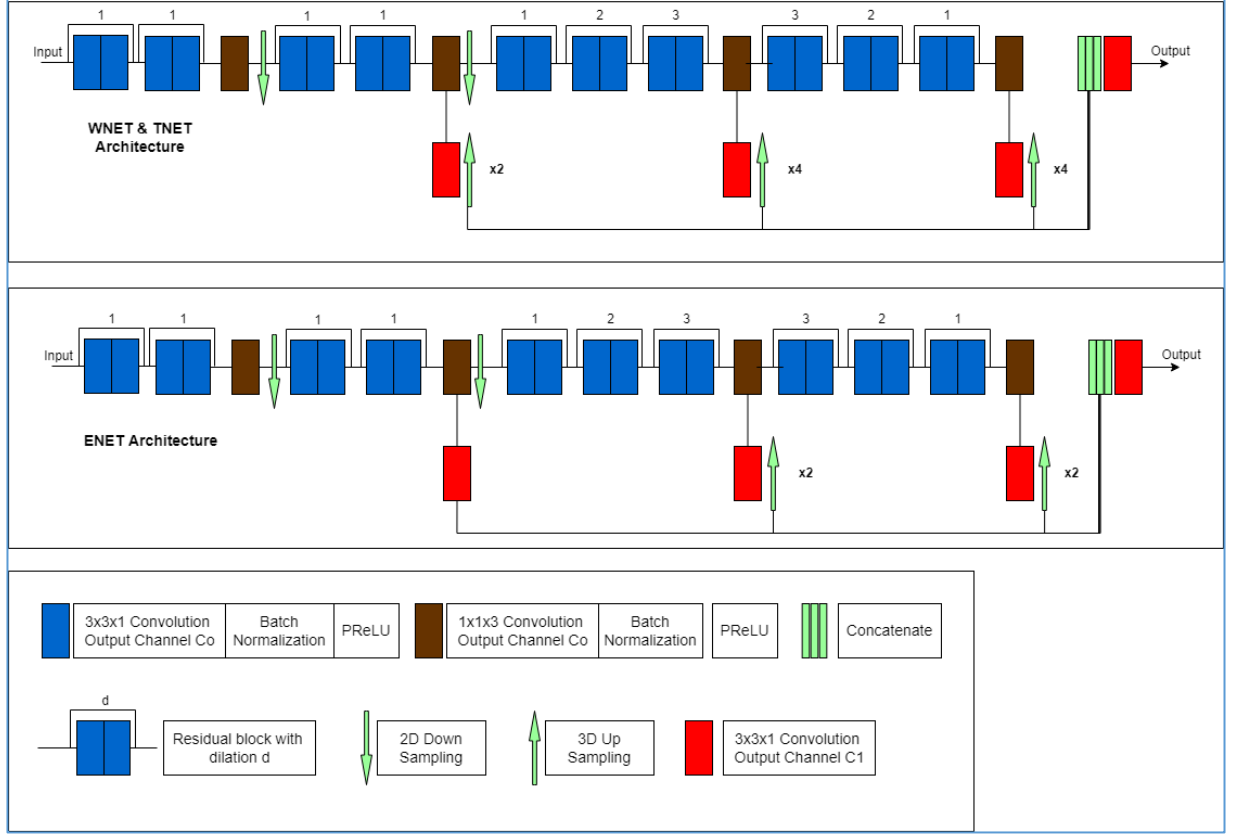


Figure 4. 3: Cascaded Anisotropic Convolutional Networks Architecture

As a trade-off, we propose anisotropic networks with a large receptive field in 2D and a small receptive field in the out-plane direction orthogonal to the 2D slices as input. WNet, TNet, and ENet have 2D receptive fields of 217×217 , 217×217 , and 113×113 , respectively. The 2D sizes of the inputs during training and testing are typically smaller than the corresponding 2D receptive fields. The out-plane receptive field of WNet, TNet, and ENet is 9. Fig.4.3 depicts the architectures of these proposed networks. They are all fully convolutional, with ten residual connection blocks that use anisotropic convolution, dilated convolution, and multi-scale prediction.

4.4.1.3 Anisotropic and Dilated Convolution Layer

To deal with anisotropic receptive fields, a 3D kernel was decomposed into an intra-slice kernel and an inter-slice kernel. Convolution layers with either of these kernels have C_0 output channels and are followed by a batch normalization layer and an activation layer, as shown in blue and brown blocks in Fig.4.3. The activation layers employ Parametric Rectified Linear Units (PReLU), which have been shown to outperform traditional rectified units. WNet and TNet employ 20 intra-slice convolution layers, four inter-slice convolution layers, and two 2D down sampling layers. ENet employs the same convolution layers as WNet, but only one down sampling layer due to its smaller input size.

Only up to two layers of down sampling were utilized to prevent significant picture quality decrease and the loss of segmentation features. Dilated convolution was used for intra-slice kernels after the down sampling layers to enlarge the receptive field within a slice. As shown in Fig.4.3, the dilation parameter was set from 1 to 3.

4.4.1.4 Residual Connection Layer

Residual connections were introduced to create identity mapping connections to bypass the parameterized layers in a network for practical deep CNN training. Ten residual blocks exist in each of our WNet, TNet, and ENet. Each of these blocks has two intra-slice convolution layers. The input of a residual block is directly added to the output, encouraging the block to learn residual functions using the input as a reference. This can smooth information propagation and accelerate training convergence.

4.4.1.5 Multi-scale Predictions

In a cascaded network, shallow layers learn to represent local and low-level features, while deep layers learn to represent higher-level, more global information. In order to mix features at various scales, we utilize three $3 \times 3 \times 1$ convolution layers at various network depths to obtain numerous intermediate predictions and up sample them to the input resolution. To obtain the final score map, a concatenation of these predictions was fed into an additional $3 \times 3 \times 1$ convolution layer. These layers are given by red blocks in Fig.4.3. The outputs of these layers have C_l number of channels, where C_l is the number of classes for segmentation in each network. C_l equals to 2 in the used method.

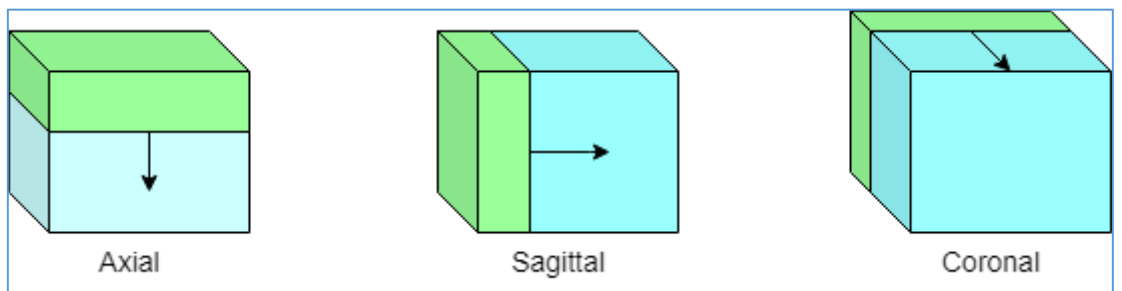


Figure 4. 4: Illustration of multi-view fusion at one level of the proposed cascade.

Due to the anisotropic receptive field of our networks, we average the SoftMax outputs in axial, sagittal and coronal views. The orange boxes show examples of sliding windows for testing. Multi-view fusion is implemented for WNet, TNet, and ENet, respectively.

4.4.1.6 Multi View Fusion

Here the segmentation results are combined from three separate orthogonal viewpoints to benefit from 3D contextual information because the anisotropic convolution has a limited receptive field in the out-plane direction. Each of the three networks—WNet, TNet, and ENet—was trained in axial, sagittal, and coronal perspectives. Predictions from these three viewpoints were used during testing to produce the final segmentation. For each level in the cascade of WNet, TNet, and ENet, we average the SoftMax outputs in these three views for the fusion. Fig.4.4 provides an example of multi-view fusion at a single level [33].

4.4.2 3D U-Net

In medical imaging, the U-Net is the most commonly used architecture for segmentation. Because of the nature of U-Net architecture, it collects low-level features that are responsible for collecting information about the shape of each class and high-level features that are responsible for deciding which class each position belongs to based on the collected information. Furthermore, it employs both (low-level and high-level features) to restore the original input's spatial dimension, with each location classified to the appropriate class.

Because 3D U-Net employs 3D convolution layers, it collects information not only along the spatial domain but also along the channel domain [34]. When using 3D convolution layers, there is a trade-off. On the one hand, they can gather information in 3D and connect it from different channel layers. However, they require more computation power and memory space than 2D convolution layers. The input and middle tensors, like in 3D convolution, will be 5D tensors with the shape $\text{Batch Size} \times \text{Channels} \times \text{Height} \times \text{Width} \times \text{Depth}$. The input and middle tensors for 2D convolutions, on the other hand, will be 4D tensors with the shape $\text{Batch Size} \times \text{Channels} \times \text{Height} \times \text{Width}$.

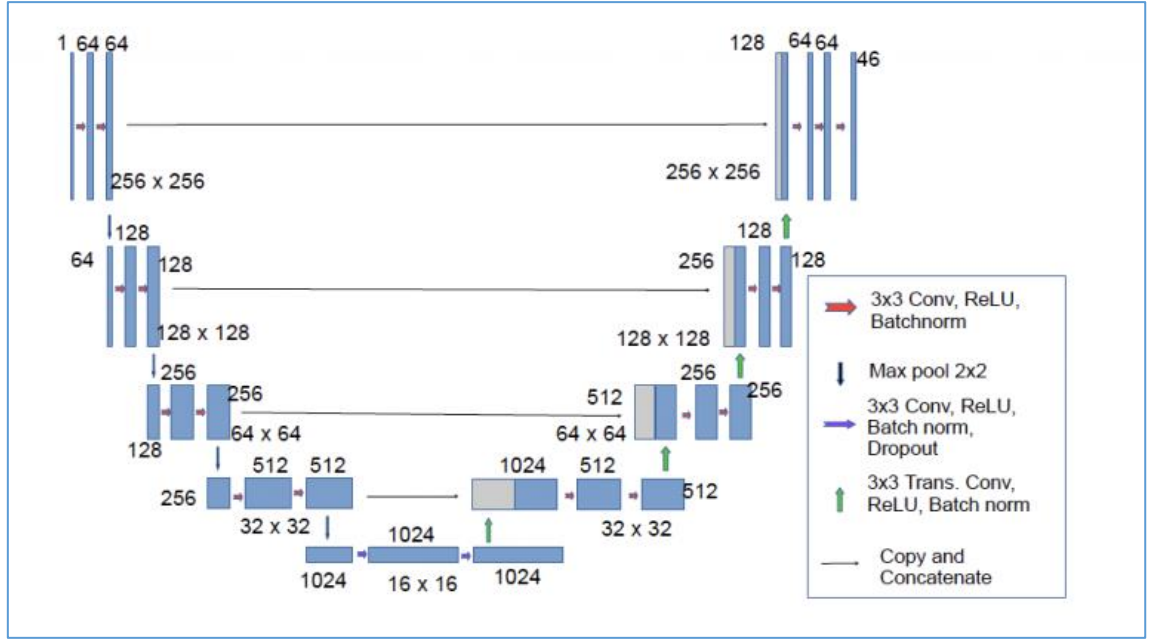


Figure 4. 5: Proposed 3D U-Net Architecture

The proposed densely connected encoder-decoder U-Net architecture for brain tumor segmentation, uses dense connections while enhancing the maximum feature size to 64 in the final output layer. The proposed U-Net architecture is depicted in Fig.4.5.

The output features at the encoder path levels are 64, 128, 256, 512, and 1024. The proposed work is divided into two parts: (i) dense blocks, which are the building blocks of the encoder-decoder paths, and (ii) residual-inception blocks, which are used in the first dense block of the encoder path as well as the decoder path up sampling layers [17].

4.4.2.1 Dense Blocks

Dense connections have demonstrated exceptional accuracy in both medical and non-medical domains. This is because dense connections have the property of feature reuse, which means that the output feature maps of all previous layers are the inputs to the subsequent layers. As a result, the dense connection's feature reusability reduces network parameters while improving segmentation accuracy. Furthermore, dense connections enable multi-path flow for gradients between layers during back-propagation training and thus implicit deep supervision [35].

As a result, we use dense networks as inspiration for each dense block. Three convolution layers are present in each block of the encoder-decoder paths. Fig.4.6 depicts the first dense block of the encoder path. A residual-inception block (RIB) and the first convolution layer's output feature maps are concatenated here. The concatenated features are then passed to the dense block's first convolution layer.

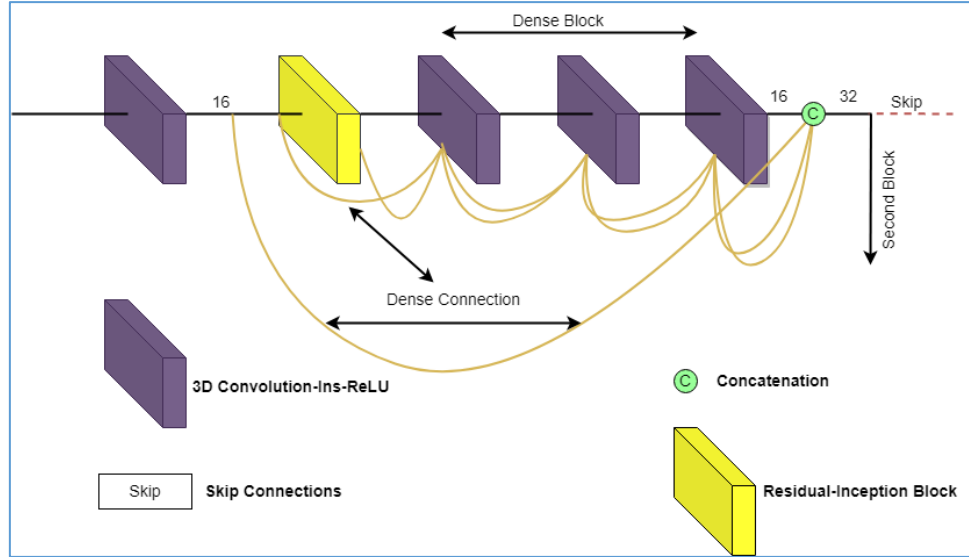


Figure 4. 6: Densely Connected Structure of the first block of encoder path

In addition, we doubled the first dense block's output-feature maps. To improve the contexts, the output-feature maps of the remaining dense blocks in the encoder path are doubled. We also employ a growth rate of 2, which aids in the reduction of training parameters and allows the proposed model to fit into GPU memory.

4.4.2.2 Residual-Inception Blocks

Three parallel dilated convolution layers comprise a Residual-Inception Block (RIB). It is used in conjunction with the encoder path's first dense block. Following that, RIB was used after each up-sampling operation in the decoder path. The RIB employs residual connections to avoid the vanishing gradient problem, while the inception blocks provide multi-scale contexts for the existing residual networks. As a result, the combined RIB architecture reduces the size of the features. Fig.4.7 depicts a RIB architecture proposed to address the problem of different tumors using different sizes of the receptive field. The RIB multi-scale contextual information can reduce the number of false positives and prevent failed segmentation.

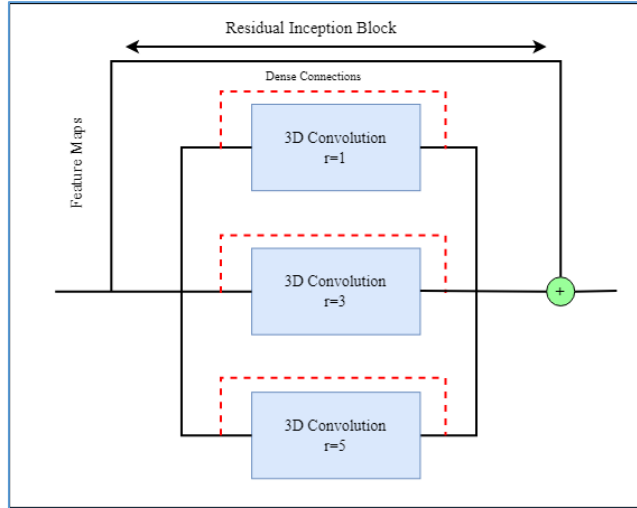


Figure 4. 7: Proposed RIB structure of the 3D UNet

Furthermore, addition operations are used for skip-connections rather than concatenation operations, which are more computationally and memory-efficient. Dense connections are kept improving the feature strength by concatenating feature maps of addition operations and the dense blocks of the decoder part. Finally, the SoftMax layer was used to achieve the desired results.

4.4.3 Pre-Trained Res-Net

Deep neural networks with many layers are used to solve complex problems. This is because these layers gradually learn complex features. When recognizing features, the first layers may learn to detect edges; the subsequent layers may learn to recognize textures, etc. However, the maximum depth threshold for a traditional convolutional neural network (CNN) limits the number of layers that can be added. Because the accuracy of traditional CNN decreases as the number of layers increases. The introduction of Res-Net or residual networks, which are made up of residual blocks, has alleviated the problem of training very deep networks [28].

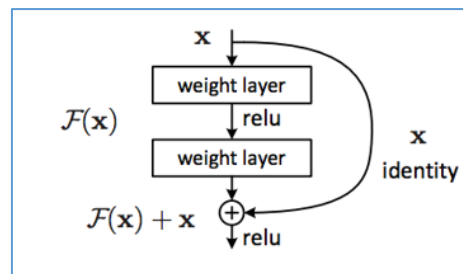


Figure 4. 8: Residual learning: a building block

4.4.3.1 Residual Connection

Fig.4.8 depicts a residual network building block. A shortcut connection and element-wise addition are used to perform the operation $(F + x)$. Batch normalization [36] is used immediately after each convolution operation and before activating the Rectified Linear Unit (ReLU). The network does not employ dropout [37]. Residual networks improve image classification performance and can be used to segment tumors.

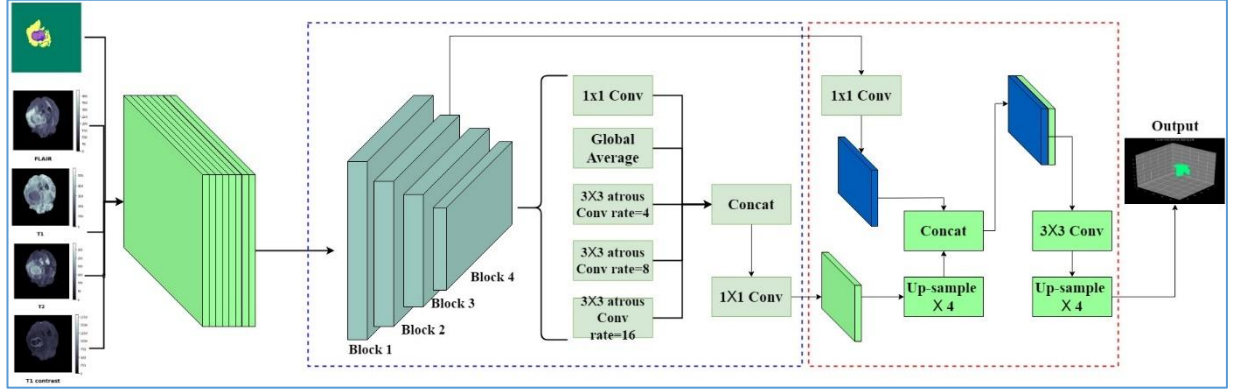


Figure 4. 9: Pre-Trained Res-Net structure.

The encoder-decoder structure has been widely used in deep neural networks for semantic segmentation in recent years. The encoder enables the network to generate high-dimensional image feature maps quickly, while the decoder allows the network to recover sharp object boundaries. The network is designed in accordance with Res-Net 34. Fig.4.9 depicts the structure of our proposed network.

4.4.3.2 Encoder

The spatial dimension of the feature maps is gradually reduced in the encoder using four blocks, as shown in Fig4.9. Fig.4.10 depicts the structure of each block, where BN denotes batch normalization and ReLU denotes rectified linear unit. To avoid overfitting during training, the ReLU activation function is used. In this paper, the ReLU parameter is set to 0.5. BN is a method for normalizing each layer's input and overcoming the internal covariate shift problem. Since stride is set to 2 in each block, the fourth block's output is 256 times smaller than the first block's input. The baseline also uses an atrous convolutional layer instead of a traditional convolutional layer. This enables us to use spatial pyramid pooling and cascaded modules to interpolate the multi-scale context while broadening the range of view of filters [38]. As illustrated in Fig. 4.10, the output of each block can be determined by adding its input to the output of the first atrous convolutional layer.

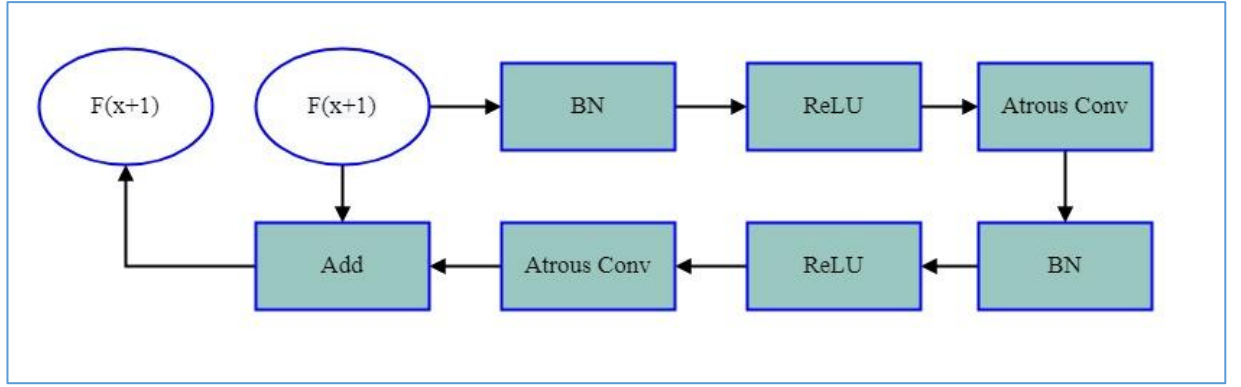


Figure 4. 10: Structure of each block unit in Fig.4.8

As shown in Fig.4.9, the output of the fourth block feeds into five branches. To obtain global image feature representations, the baseline employs global average pooling. In addition, three atrous convolutional layers with varying rates are used to collect multi-scale data. The rates are determined entirely by the feature map generated by block 4 and are set to 4, 8, and 16, respectively. Finally, the output of the five branches is concatenated and compressed using a 1×1 convolutional layer.

4.4.3.3 Decoder

The baseline in the decoder applies to skip connection to the feature map produced by the second block. As a result, the high-level feature map's local feature details are substantially enhanced. After that, the high-level and low-level feature maps are integrated. After a feature map up-sampling process, a probability map can be obtained. The semantic segmentation result can be obtained by locating pixels with probabilities more significant than our defined threshold.

The implemented PT Res-Net model is a typical Res-Net model similar to the encoding-decoding model. The encoding and decoding blocks are similar to the encoder-decoder model. Instead of the max-pooling layer, a custom layer termed as Max-Pooling with Argmax is used, which performs typical max-pooling on the feature maps with a stride of 2×2 and kernel size of 2×2 but also stores the index position of the max-pooling pixels. Similarly, instead of the up-sampling layer, a custom layer termed Max Up Sampling with Argmax is used, which uses the corresponding index values stored during pooling to up-sample the feature maps.

The final output layer of all the models is a convolutional layer of kernel size 1×1 and the number of feature maps is equal to the number of classes and is followed by a sigmoid activation function to get a probability feature map.

Skip-connections and residual elements address the localization aspect of semantic segmentation networks by allowing low-level representations to pass through the network and informing the latest layers about fine-grained spatial details.

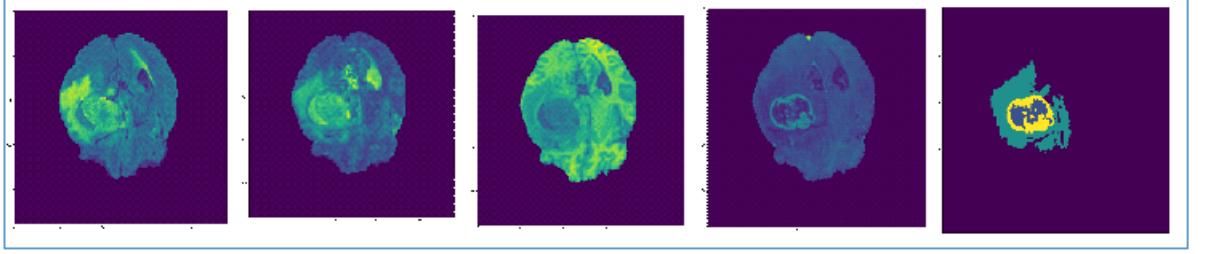


Figure 4. 11: Axial representation of an example patch of size 80 x 80 x 80.

However, because these networks are focused on proper boundary alignment, the classification aspect of semantic segmentation networks, which deals with the proper identification of the delineated structures, is hampered. Modern classification architectures rely on globally connected layers, which correspond to a fully connected layer in the most extreme case (all nodes are connected). It was obvious that such an operation is not possible in a fully convolutional architecture; however, we can approximate global connectivity in convolutions by increasing kernel size: in the limit, the kernel is as large as the input feature map, which can be interpreted as a fully connected layer.

4.4.4 Ensemble Model

To improve the performance of the segmentation and reduce the proposed model architectures' variance and use a voting strategy (majority vote) to build an ensemble model without a weighted schema. During the training process, different models were trained individually. The choice of the number of iterations for the training process was based on the model's performance in the validation set. In the testing phase, each model predicts each voxel's class individually, and the final class is determined by a majority vote [39].

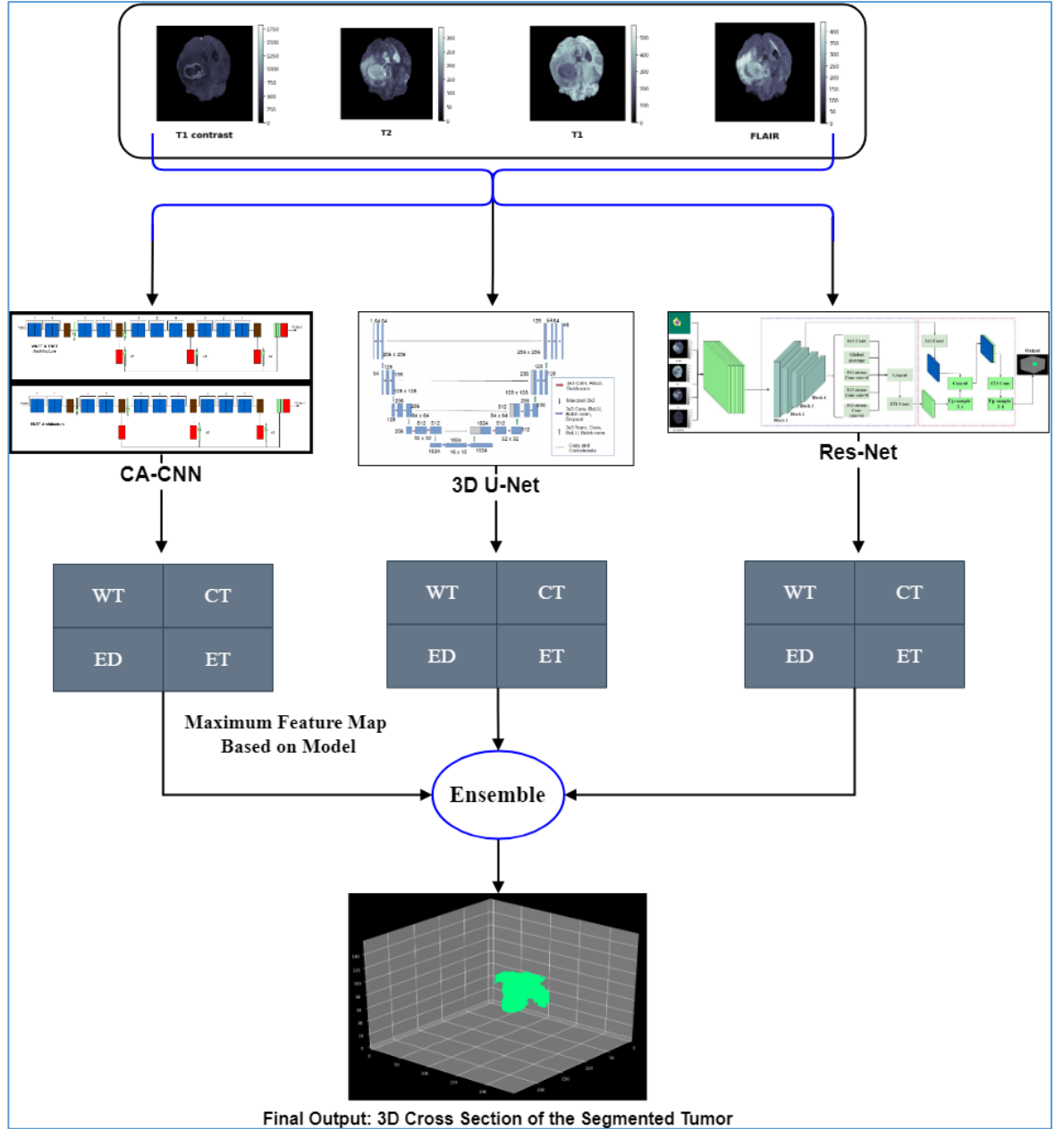


Figure 4. 12: Proposed Ensemble Model for Tumor Segmentation

4.5 Implementation

Pytorch has been the Deep Learning framework of choice, and all code has been written on top of it. It includes a comprehensive set of libraries and utilities for training and serving Machine Learning models, particularly Deep Learning models and, more specifically, CNNs. It has a large developer community and evolves quickly, allowing developers to experiment with the most recent advances in the field. We make extensive use of its high-level APIs.

The former simplifies the development of ETL processes such as the one described in the section Pipeline for Data Ingestion. In contrast, the latter accelerates the design, implementation, training, and evaluation of CNN architectures such as those described in the section Network Architectures above. One of the primary benefits of using such high-level APIs over low-level approaches is that they enable developers to easily integrate new and powerful features such as Distributed training or GPU support, which speeds up the development experimentation serving cycle. However, because they are primarily focused on more common domains such as computer vision with realistic images or natural language processing, these APIs require extensive customization to be properly integrated in the context of medical data processing.

4.6 Experimental Design

4.6.1 Dataset Splits

The BraTS dataset is divided into two parts: training and validation. Test data is available from multiple sources, including the test set used for ranking in the BraTS challenge and the clinical data stored by the FastMRI dataset. This eliminates the need for the third split for the test. We use a split ratio of 75 - 25% for training and validation. The BraTS dataset results in 369 subjects for training and 125 subjects for evaluation.

4.6.2 Evaluation

Model evaluation targets tumoral structures that combine a subset of intra-tumoral classes. In particular, we assess several measures on the following structures, as detailed further in this section.

- **Whole Tumor(WT)**: Includes all tumor structures.
- **Enhancing Tumor(ET)**: Includes only the class of enhancing tumor.
- **Tumor Core(TC)**: Includes the Necrotic/Non-enhancing tumor and enhancing tumor structures except for the edema region.

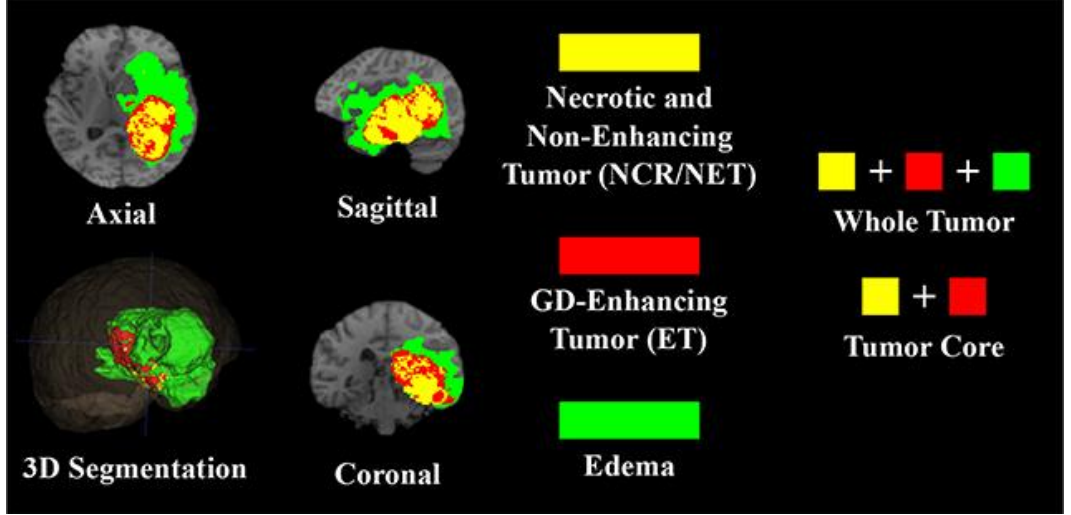


Figure 4. 13: Tumoral structures of the 3D brain tumor.

All measurements used to evaluate the model are based on four basic measures of the confusion matrix.

- **True Positives (TP):** Positive examples that have been predicted to be true.
- **True Negatives (TN):** Negative examples that have been predicted as true.
- **False Positives (FP):** False examples that have been predicted as positive.
- **False Negative (FN):** True examples that have been predicted as negative

4.6.3 Activation Function and Loss Function

One of the most important learning parameters for any deep learning model is the activation function. The activation function governs the output, accuracy, and computational efficiency of a model's training. All of the models in our methodology use the ReLU and sigmoid activation functions. The ReLU function (equation (3)) is defined as:

$$f(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (3)$$

The sigmoid function (equation (4)) is defined as:

$$f(x) = \frac{1}{1 + \sqrt{e^{-x}}} \quad (4)$$

The primary evaluation score is the Dice coefficient, usually abbreviated as DICE or Dice Similarity Coefficient (DSC), which is a statistic used for comparing the similarity of two sets A and B:

$$DICE(A, B) = \frac{2|A \cap B|}{|A| + |B|} \quad (5)$$

Here, the two sets, A and B, that are compared in this formula are the set of true examples and the set of positive examples. Therefore, the formula can be re-written as:

$$DICE\ Loss = \frac{2TP}{2TP + FP + FN} \quad (6)$$

Jaccard index, also known as Jaccard Similarity Coefficient or Intersection Over Union (IoU), is a statistic used in gauging the similarity and diversity of sample sets in the process of evaluation. Formula for two sets, A and B, is given as:

$$IoU(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (7)$$

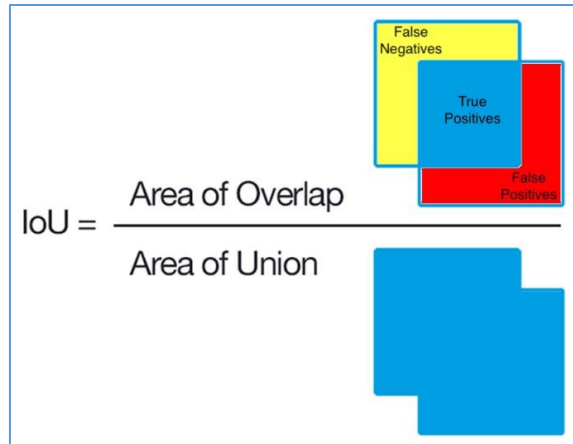


Figure 4. 14: Jaccard Index depicted with an image representation

Other metrics necessitate pixel-by-pixel comparisons of the ground truth and predicted masks. True Positive (TP) is the total number of positive pixels (belonging to the tumor) in the ground truth that are correctly predicted as a tumor, True Negative (TN) is the total number of negative pixels (belonging to the background) in the ground truth that are correctly predicted negative, False Positive (FP) is the total number of negative pixels that are falsely predicted as positive pixels, and False Negative (FN) is the total number of negative pixels that are falsely predicted as positive pixels.

4.6.4 Training Details

The loss function to be minimized during training was weighted categorical cross-entropy. To extend it to 4D tensors, a custom implementation of such loss was required. Weights for each class could be specified to explicitly penalize specific classes. The penalty ratios for L1 and L2 penalization of the weights (for regularization purposes) were $(1e - 6)$ and $(1e - 4)$, respectively, for the entire loss function to be minimized.

The ADAM optimizer [40] was used to train all of the models. In all experiments, the initial learning rate was set to $(1e - 3)$, and a learning rate decay policy was implemented to stabilize training as the training procedure progressed. In the experimental set exponential decay of the learning rate with a decay rate of 0.9 every 1000 training steps were used. The number of training steps ranged from 30000 to 75000, depending on the model's complexity and the dataset used. The training procedure alternated 1000 training steps with one complete model evaluation.

Batch size was also variable, ranging from 4 to 12 depending on the number of network parameters and memory requirements; in all cases, we aimed to maximize batch size within the memory limits of the available hardware for efficiency. During training, the data ingestion pipeline was set to extract patches of size 80x80x80 with a 50% chance of being centered on a background voxel and a 50% chance of being centered on a tumor voxel (50 percent background, 20 percent edema, 15 percent enhancing tumor and 15 percent necrosis and non-enhancing tumor). During the evaluation, preprocessed whole brain volumes were used to provide a realistic value of performance in a real-world scenario requiring whole brain volume inference.

As we discussed here, the training will be done on the model architectures separately and the final result will be taken through the ensemble model.

4.6.5 Modality Combinations

Data ablation experiments are conducted by limiting the available input modalities at training time while always maintaining the bare minimum of modalities required to correctly identify all structures, namely T1-Gd and FLAIR. Such experiments are motivated by two factors. First, we want to evaluate each modality's relative contribution to the overall segmentation and determine whether some modalities are redundant or provide helpful information.

Second, having models that can work with a limited number of modalities is convenient, if not necessary (in some clinical cases, not all MR sequences are included in the protocol), even if such models with restricted input information do not perform as well as models trained without data restrictions.

4.7 Testing Environment

4.7.1 Software specification

- Operating Environment: Google Colab Pro

4.7.2 Hosted Hardware specifications

- Processor: Intel Core i7
- RAM: 35GB
- Disk Space: 225 GB
- GPU: Python 3 Google Compute Engine backend (TPU)
- GPU RAM: 35 GB

4.7.3 Programming Technologies

- Python 3.9
- Pytorch 1.2
- CUDA 10.0
- Numpy
- Scikit-learn
- Pydicom
- Pandas

CHAPTER 05

RESULTS AND DISCUSSION

5.1 Overview of the Results and Discussion

This chapter highlights the results of this study along with segmented tumor images. The segmentation results of each model used, and the ensemble model are presented separately. Every result is followed by a discussion of how they have worked out. Results will be discussed relatively to Fig.5.1 which shows the workflow of this study.

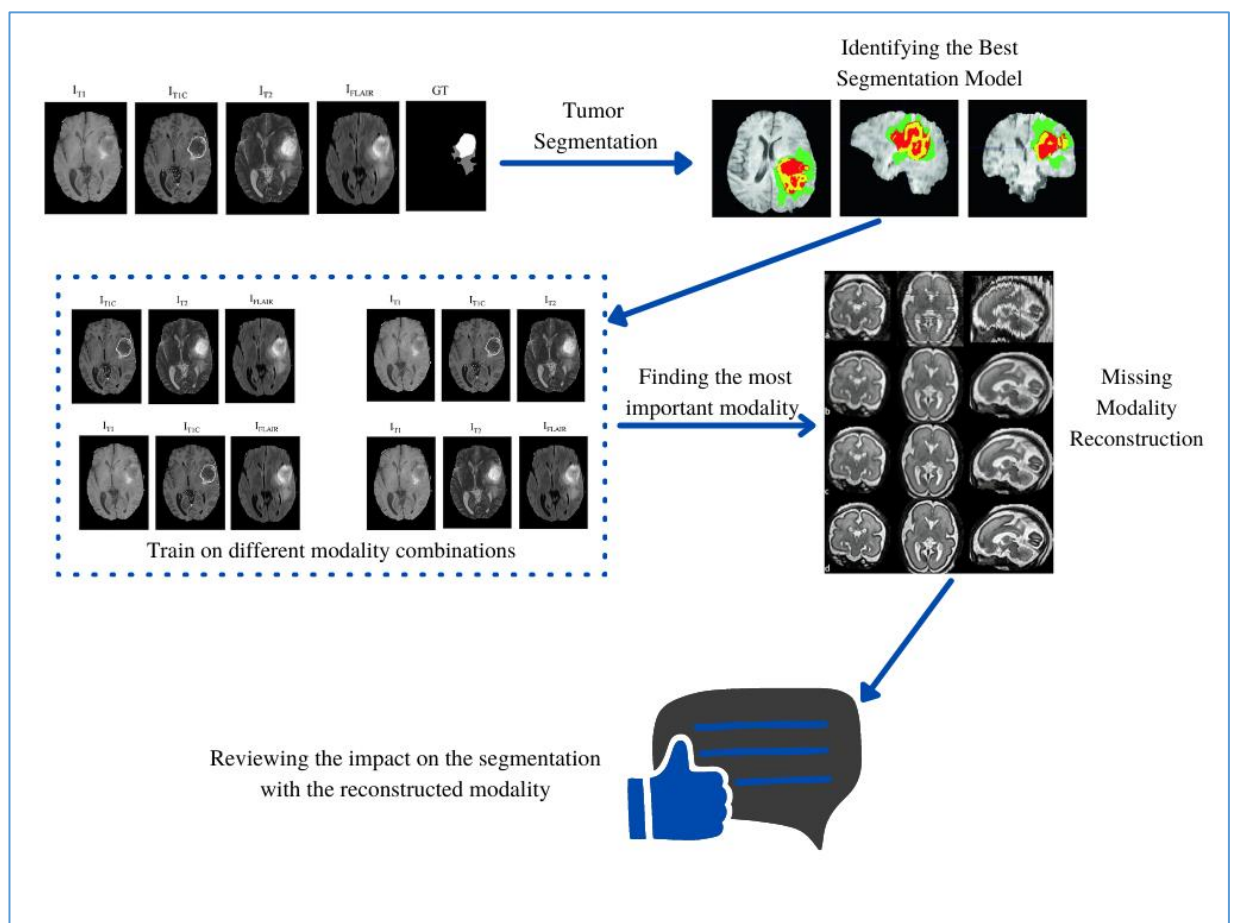


Figure 5. 1: Results Workflow

5.2 Segmentation Results on all the CNN Architectures

Table 5. 1: Segmentation results on every model architecture

Model	Metric	Enhancing Tumor (ET)	Whole Tumor (WT)	Tumor Core (TC)
CA-CNN	Mean DICE	0.72682	0.60282	0.85392
	Sensitivity	0.81258	0.93045	0.85305
	Specificity	0.99807	0.99336	0.99786
3D U-NET	Mean DICE	0.72088	0.88762	0.82567
	Sensitivity	0.84281	0.90188	0.81913
	Specificity	0.99743	0.99416	0.99813
PT-RES-NET	Mean DICE	0.70524	0.86485	0.82531
	Sensitivity	0.87452	0.88864	0.79658
	Specificity	0.83243	0.91056	0.91069
ENSEMBLE	Mean DICE	0.74687	0.88152	0.78343
	Sensitivity	0.83064	0.90688	0.83156
	Specificity	0.99815	0.99549	0.99863

From the results in table 5.1 we can see that Cascaded CNN and the Ensemble model architectures gives higher segmentation accuracies on the whole tumor, enhancing tumor and the tumor core regions. The study presents the DICE coefficients for all target modalities of all the architectures, trained with different data configurations.

When training with the BraTS dataset, all the training and validation subjects are considered for the whole tumor, enhancing tumor and tumor core statistics. From table 5.1, we can observe that the differences among the methods are minimal, apart from U-Net. However, the best model is the Ensemble model trained with all the available MRI modalities, as it presents the best DICE scores in the validation set for the whole tumor and enhanced tumor structures.

However, the Cascaded CNN performs similarly, and it is even better at delineating and identifying the tumor core. It is particularly remarkable that the best model for predicting the tumor core is the Cascaded CNN trained with T1-Gd, T2, FLAIR and T1 modalities. This model performs worse at whole tumor segmentation than Pre-trained Res-Net or 3D U-Net trained with all modalities, but it outperforms them when segmenting the enhancing tumor and the tumor core. Such behavior can be attributed to the exclusion of T2's redundant information, as FLAIR provides almost the same information about edema, encouraging the network to pay special attention to other modalities and consequently other structures.

The performance of the models at segmenting tumor core and enhancing tumor is considerably below the segmentation performance of whole tumor. However, from a clinical perspective, it is more relevant to properly segment the whole tumor than any other structure. A radiologist can focus on precisely delineating structures such as the enhancing core if he or she is provided with an accurate segmentation of the whole tumor but it is much more complicated for him or her to assess the extent of the tumor given only core tumoral structures.

5.3 Modality Combination Experimentation on the Ensemble

In this experimental setup, there will be training and validation results on the dataset with different MRI modality combinations. These experiments are carried out on the ensemble model as it suits to be the best choice for brain tumor segmentation among the models used in this research study. All the graphs under this section represents the y-axis for total loss and the x-axis for no of batch cycles.

5.3.1 Segmentation Results on All four modalities

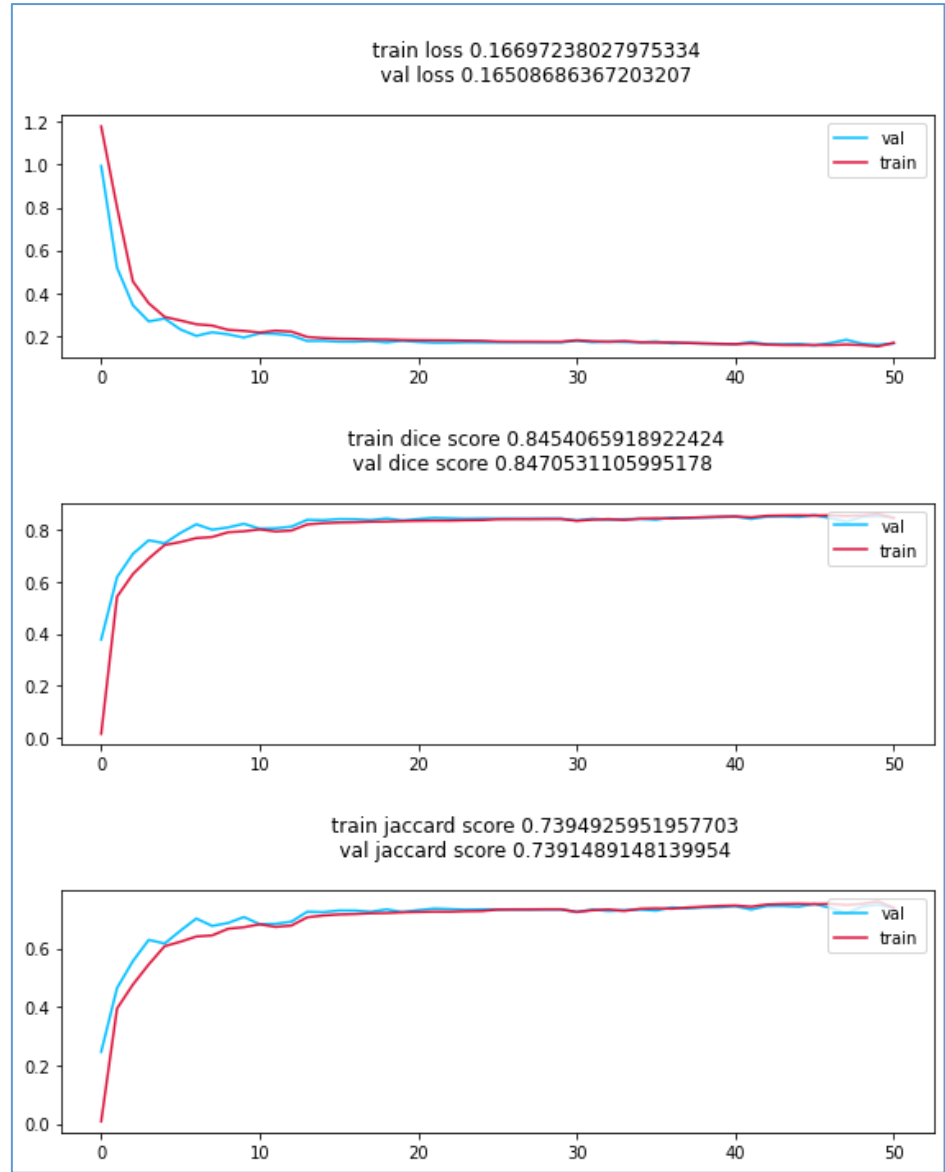
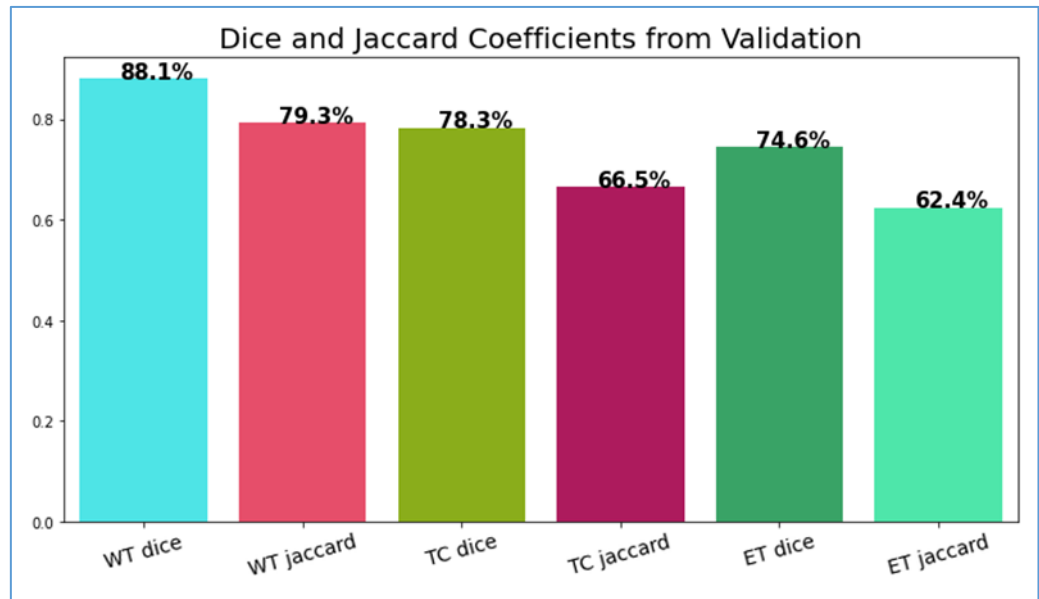


Figure 5. 2: Evaluation graphs on all the modalities

Table 5. 2: Segmentation Results on all 4 modalities

	Training	Validation
Loss	0.1669723	0.16508686
Mean Dice	0.8454065	0.84705311
Jaccard	0.7394925	0.73914891

As you can see in the table 5.2 the training loss seems to be less than 0.5 which is a good sign of better training on the segmentation. And the Mean Dice(F1-score) score gives an 84% in both training and validation which interprets a higher performance in brain tumor segmentation with regard to all 4 modalities.

**Figure 5. 3:** Dice and Jaccard from Validation for all modalities

As given in the bar chart in Fig.5.3, the Dice and Jaccard coefficients of tumor sub regions, we can see that the whole tumor segmentation performs to be with the highest score. 88.1% Dice score and 79.3% Jaccard score on the whole tumor which interprets that all four modalities provide features for the whole tumor segmentation compared to Tumor core and Enhancing tumor.

5.3.2 Segmentation Results of FLAIR, T1, and T1-ce Modalities with the Ensemble

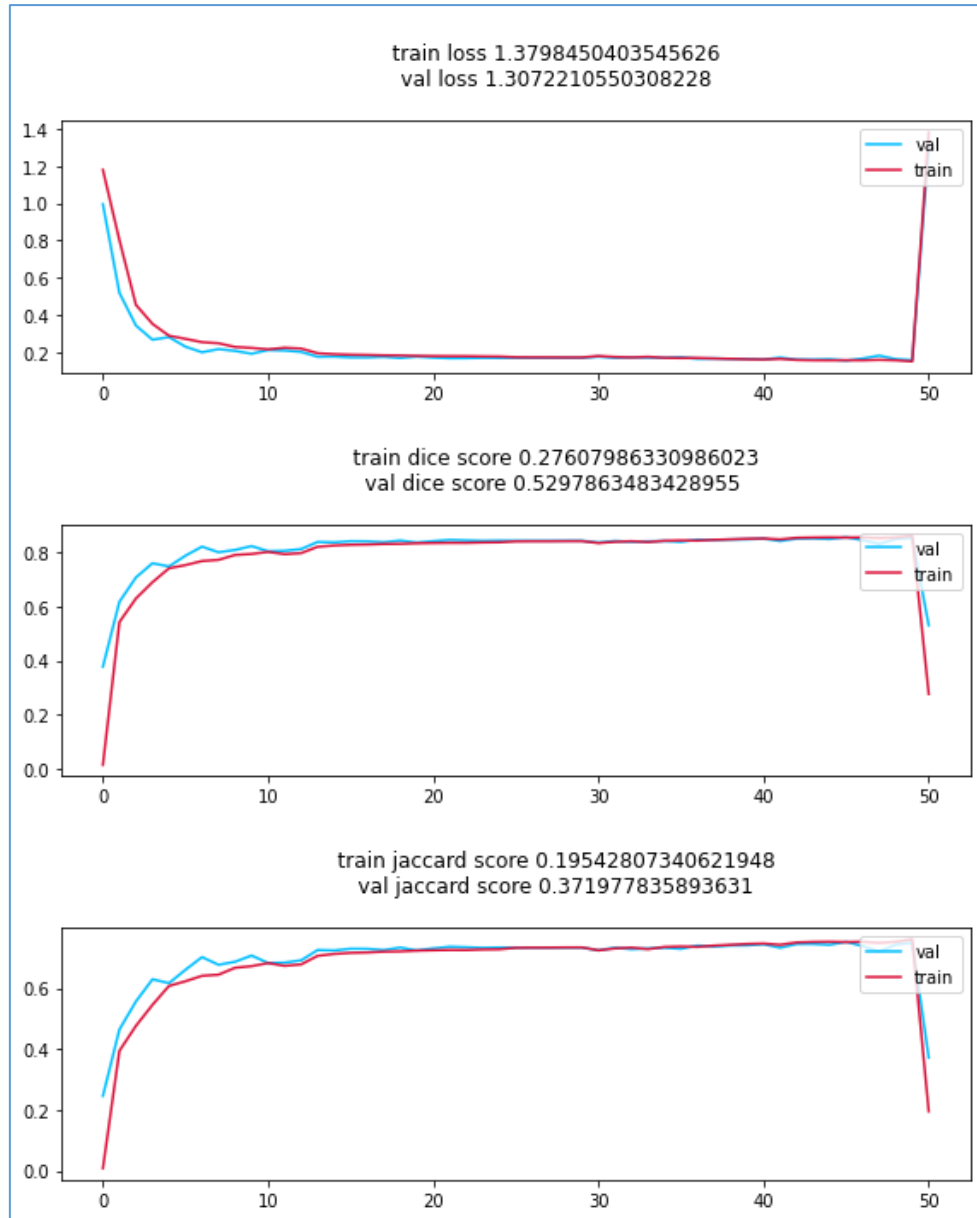


Figure 5. 4: Evaluation graphs on FLAIR, T1, and T1-ce

Table 5. 3: Segmentation Results on FLAIR, T1, and T1-ce Combination

	Training	Validation
Loss	1.3798450	1.30722105
Mean Dice	0.2760798	0.52978634
Jaccard	0.1954280	0.37197783

In the table 5.3, training loss seems to be greater than 1 which is a sign of poor training on the segmentation. And the Mean Dice(F1-score) score gives an 27% in training and 53% in validation which interprets a poor performance in brain tumor segmentation with regard to the used modality combination. Mean Jaccard score is even poor with this combination. So, the segmentation seems to lose more features from the T2 modality as the scores are lower compared to all four modalities.

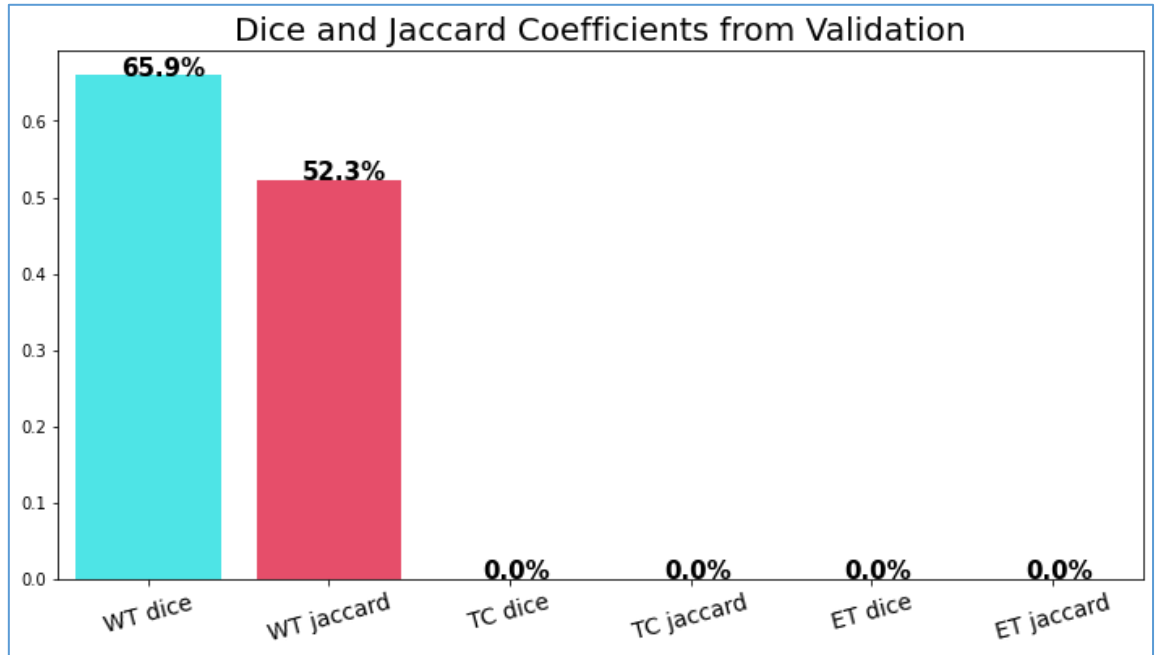


Figure 5. 5: Dice and Jaccard from Validation for FLAIR, T1, and T1-ce

As given in the bar chart in Fig.5.5, the Dice and Jaccard coefficients of tumor sub regions, we can see that only the whole tumor segmentation is accomplished with a score of 65.9% Dice score and 52.3% Jaccard score. This interprets that T2 modality is crucial in segmenting the tumor core and Enhancing tumor along with the rest of the modalities. The comparatively higher score on the whole tumor shows that the T2 modality features are not prominent for the whole tumor segmentation.

5.3.3 Segmentation Results of FLAIR, T1, and T2 Modalities with the Ensemble

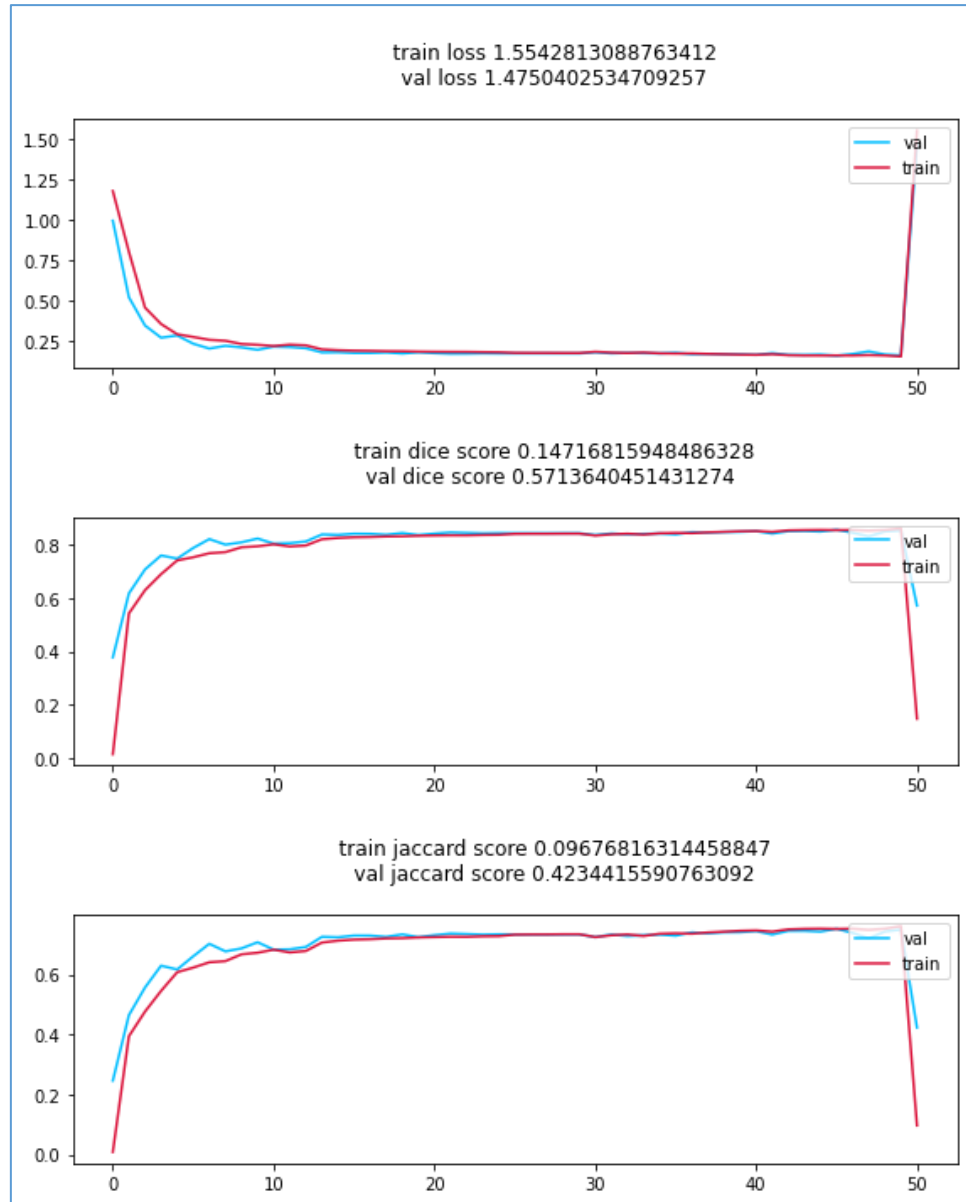


Figure 5. 6: Evaluation graphs on FLAIR, T1, and T2

Table 5. 4: Segmentation Results on FLAIR, T1, and T1-ce

	Training	Validation
Loss	1.5542813	1.47504025
Mean Dice	0.1471681	0.57136404
Jaccard	0.0967681	0.42344155

In the table 5.4, training loss seems to be greater than 1 which is a sign of poor training on the segmentation. And the Mean Dice(F1-score) score gives an 14% in training and 57% in validation which interprets a poor performance on training but comparatively higher performance in validation for the segmentation with regard to the used modality combination. Mean Jaccard score is even poor with this combination. So, the segmentation seems to lose quite less features than T2 from the T1-ce modality compared to the previous modality combination.

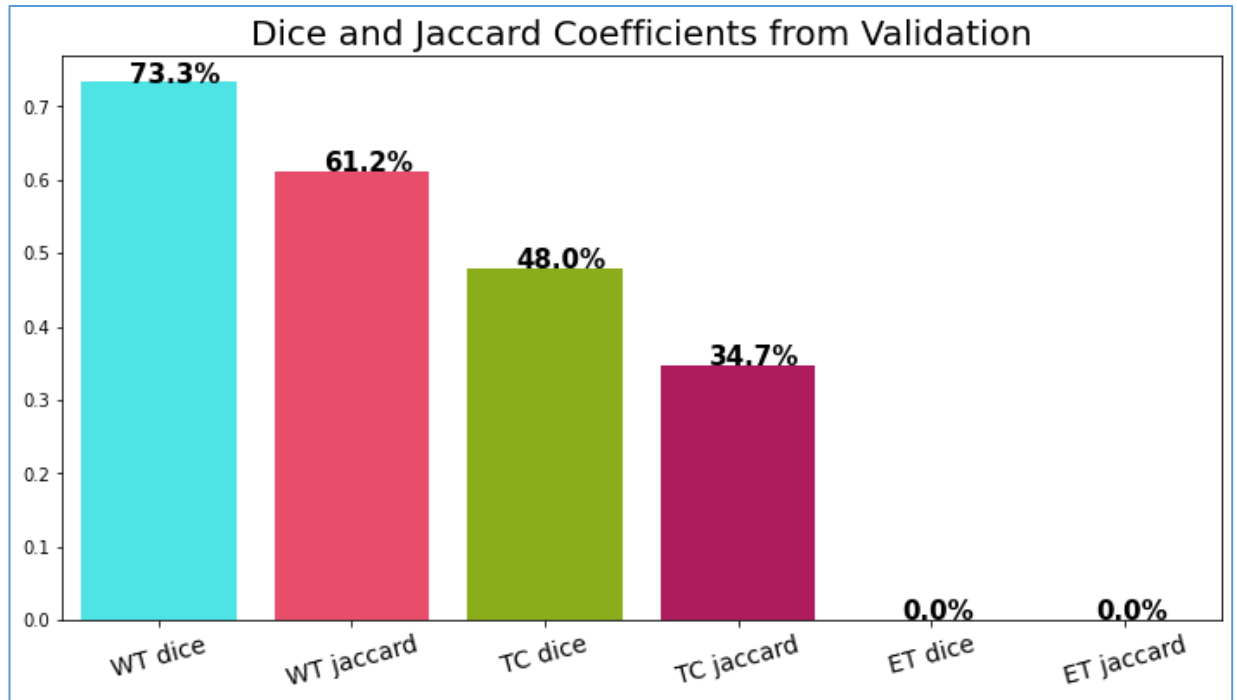


Figure 5. 7: Dice and Jaccard from Validation for FLAIR, T1, and T2

As given in the bar chart in Fig.5.7, the Dice and Jaccard coefficients of tumor sub regions, we can see that the whole tumor segmentation performs to be with the highest score and the Tumor core to have a comparatively better score. 73.3% Dice score and 62.2% Jaccard score on the whole tumor which interprets that the used modality combination provides potent features for the whole tumor segmentation and the tumor core segmentation, except for the enhancing tumor. This shows that T2 was required for the tumor core segmentation and along with T1, T2, and FLAIR modality T1-ce was important in identifying the enhancing tumor subregion

5.3.4 Segmentation Results of FLAIR, T2, and T1-ce Modalities with the Ensemble

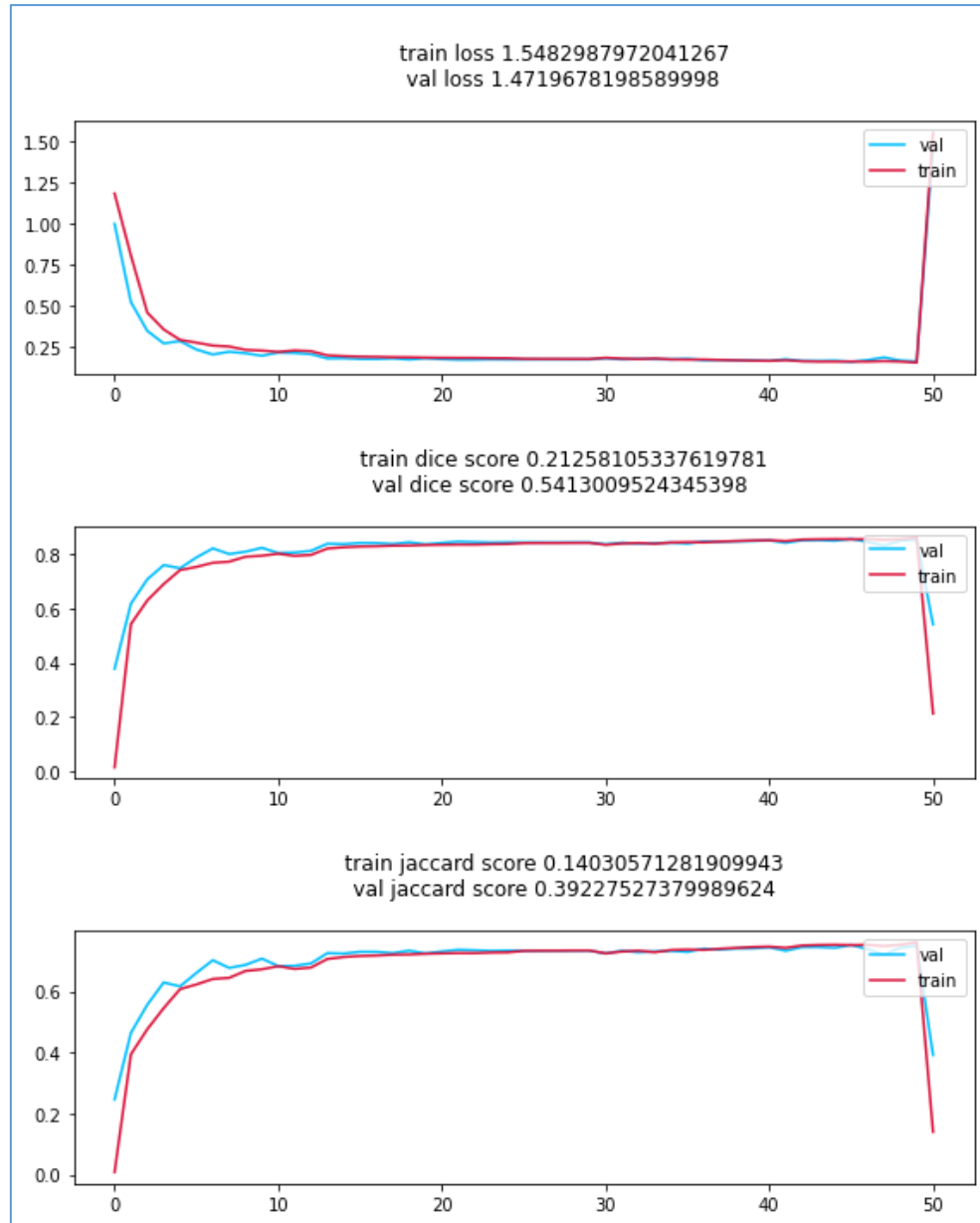


Figure 5. 8: Evaluation graphs on FLAIR, T2, and T1-ce

Table 5. 5: Segmentation Results on FLAIR, T2, and T1-ce

	Training	Validation
Loss	1.5482987	1.47196781
Mean Dice	0.2125810	0.54130095
Jaccard	0.1403057	0.39227527

In the table 5.5 above, training loss is greater than 1, which gives the idea of poor training on the segmentation. And the Mean Dice(F1-score) score gives an 21% in training and 54% in validation which interprets a poor performance on training but comparatively higher performance in validation for segmentation with regard to the used modality combination. Mean Jaccard score is even poor with this combination. So, the segmentation seems to lose more features compared to T2 and T1-ce.

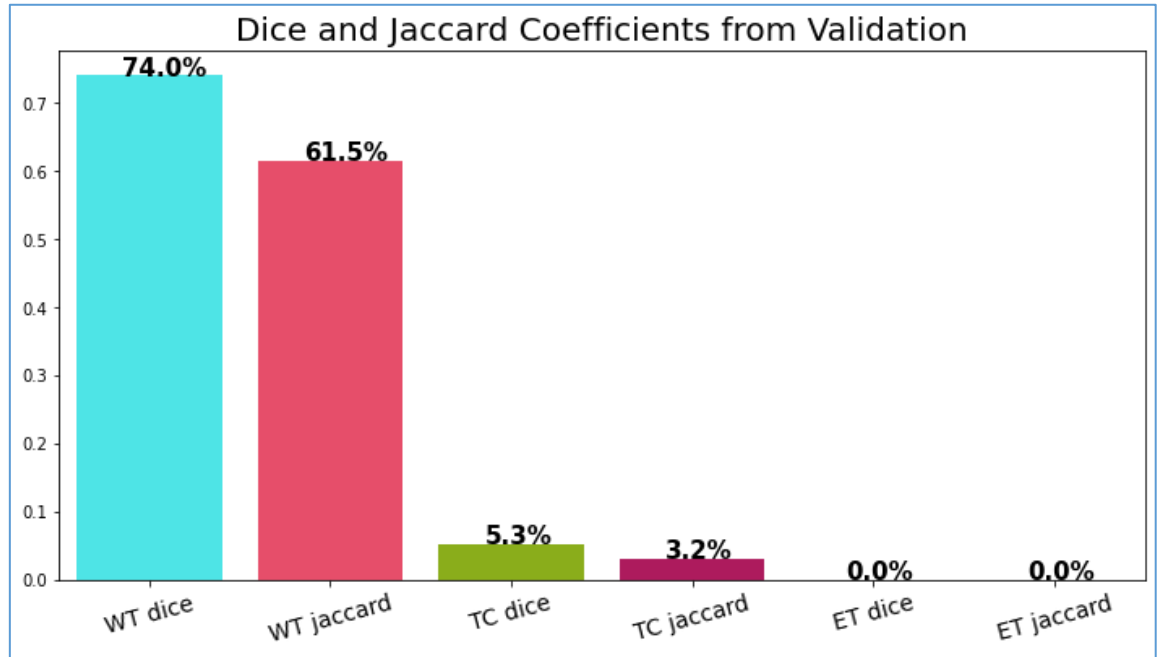


Figure 5. 2: Dice and Jaccard from Validation for FLAIR, T2, and T1-ce

As given in the bar chart in Fig.5.9, of the Dice and Jaccard coefficients of tumor sub regions, we can see that the whole tumor segmentation performs to be with the highest score and the Tumor core to have a very low score. 74% Dice score and 61.5% Jaccard score on the whole tumor which interprets that the used modality combination provides potent features for the whole tumor segmentation but not much for the tumor core segmentation, except for the enhancing tumor to have no score. This shows that T1 was required for higher score on the tumor core segmentation and along with T1-ce, T2, and FLAIR modality T1 was important in identifying the enhancing tumor subregion.

5.3.5 Segmentation Results of T1, T2, and T1-ce Modalities with the Ensemble

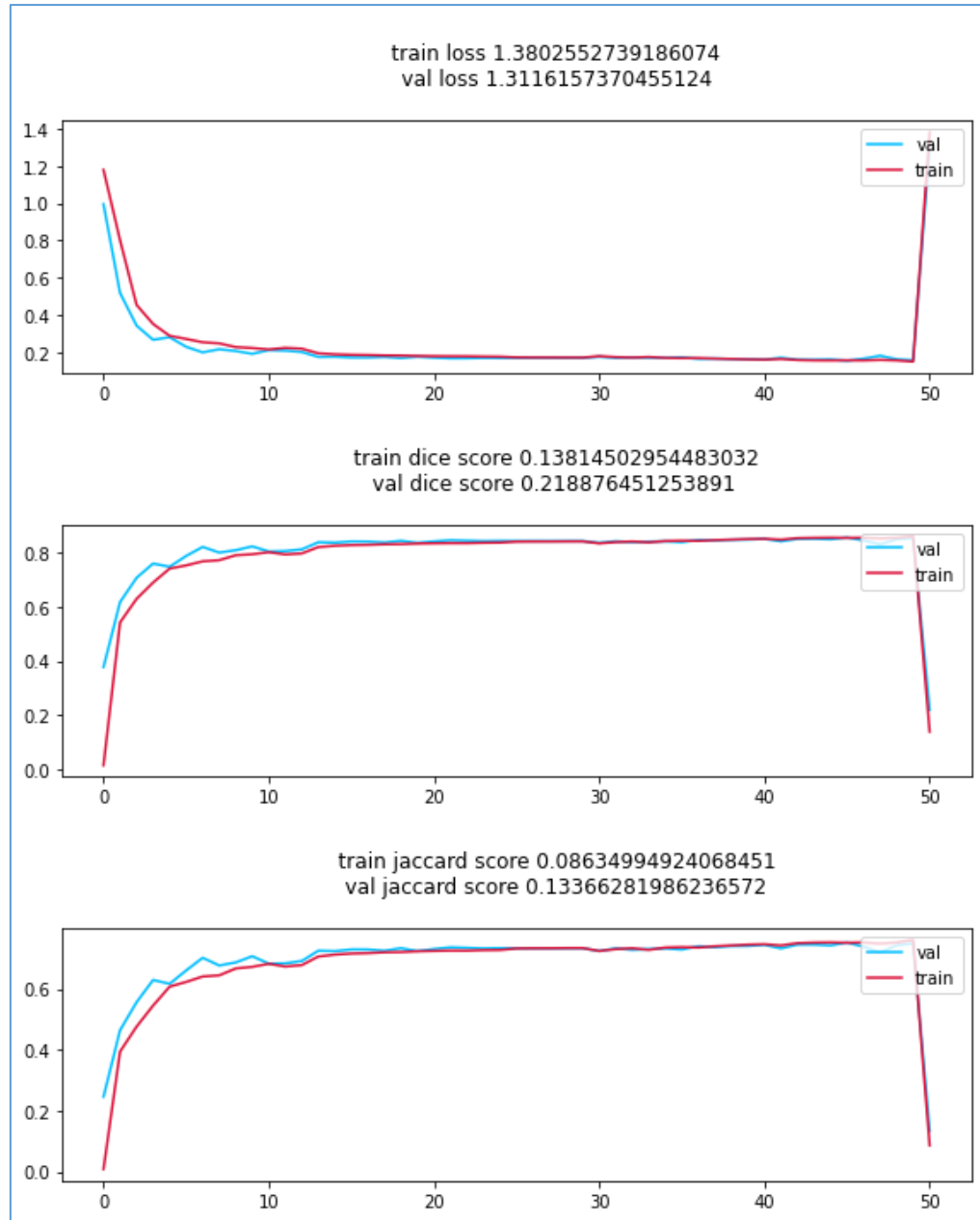


Figure 5. 10: Evaluation graphs on T2, T1, and T1-ce

Table 5. 6: Segmentation Results on T1, T2, and T1-ce

	Training	Validation
Loss	1.3802552	1.31161573
Mean Dice	0.1381450	0.21887645
Jaccard	0.0863499	0.13366281

In the table 5.6 above, training and validation loss are both greater than 1, which gives the idea of poor training on the segmentation. And the Mean Dice(F1-score) score gives an 13% in training and 21% in validation which interprets a poor performance on training but comparatively higher performance in validation for segmentation with regard to the used modality combination. Mean Jaccard score is even poor with this combination. So, the segmentation seems to lose more features without FLAIR modality.

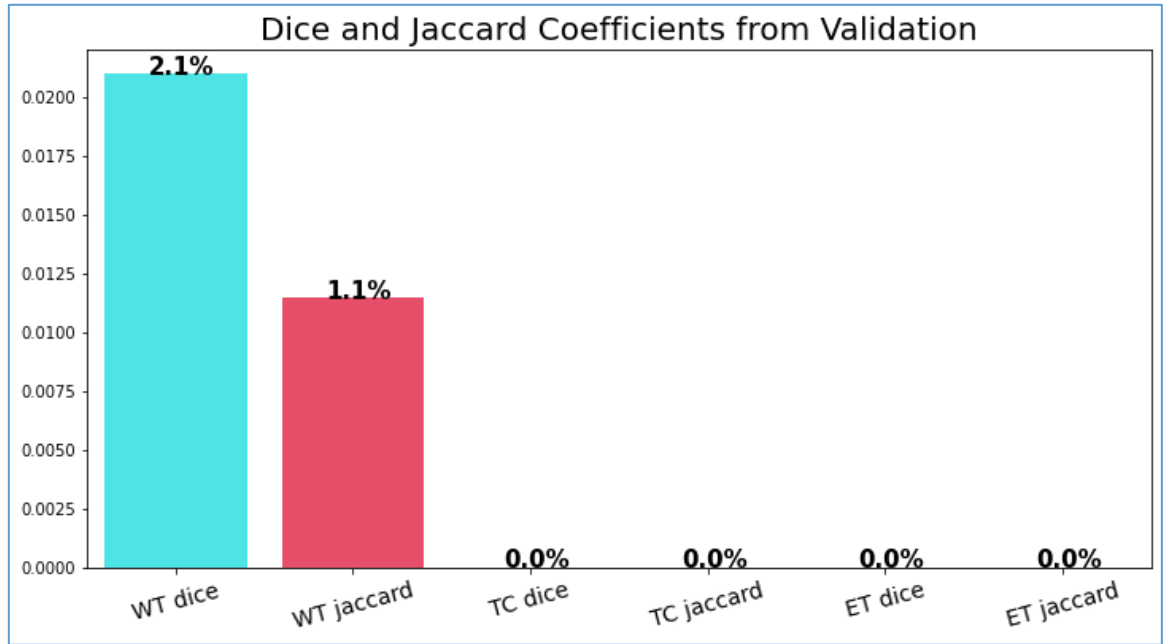


Figure 5. 11: Dice and Jaccard from Validation for T2, T1, and T1-ce

As given in the bar chart in Fig.5.11, of the Dice and Jaccard coefficients of tumor sub regions, we can see that the whole tumor segmentation performs to be very poor, and the Tumor core and Enhanced tumor have no score on segmentation. 2.1% Dice score and 1.1% Jaccard score on the whole tumor which interprets that the used modality combination does not provide potent features for the tumor segmentation at all. This shows that FLAIR modality is very important in tumor segmentation.

5.4 Predicted Segmentations vs Ground Truth

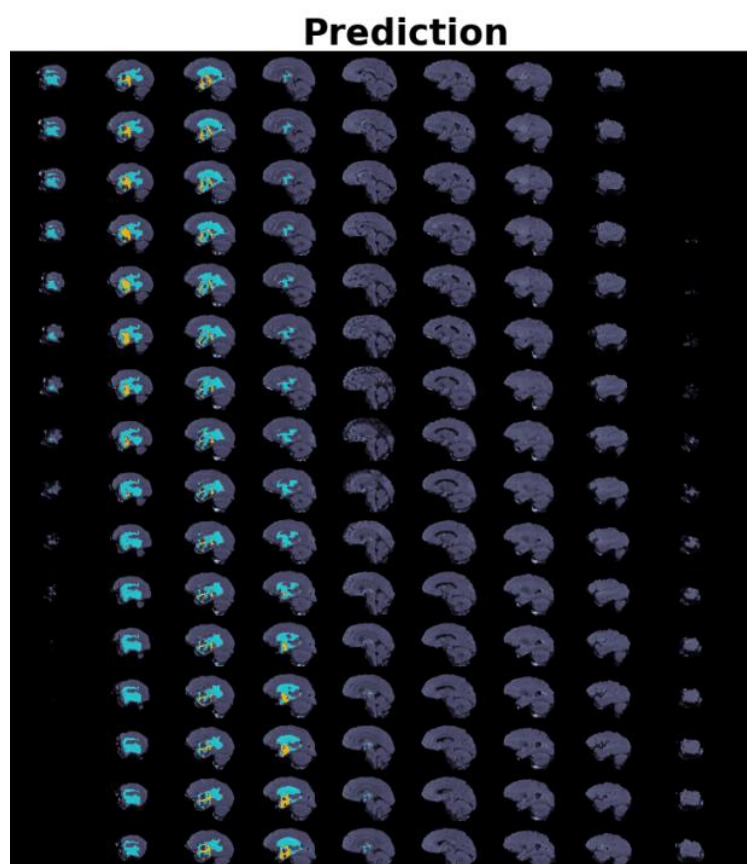


Figure 5. 123: Predicted output of a Patient's Brain tumor segmentation

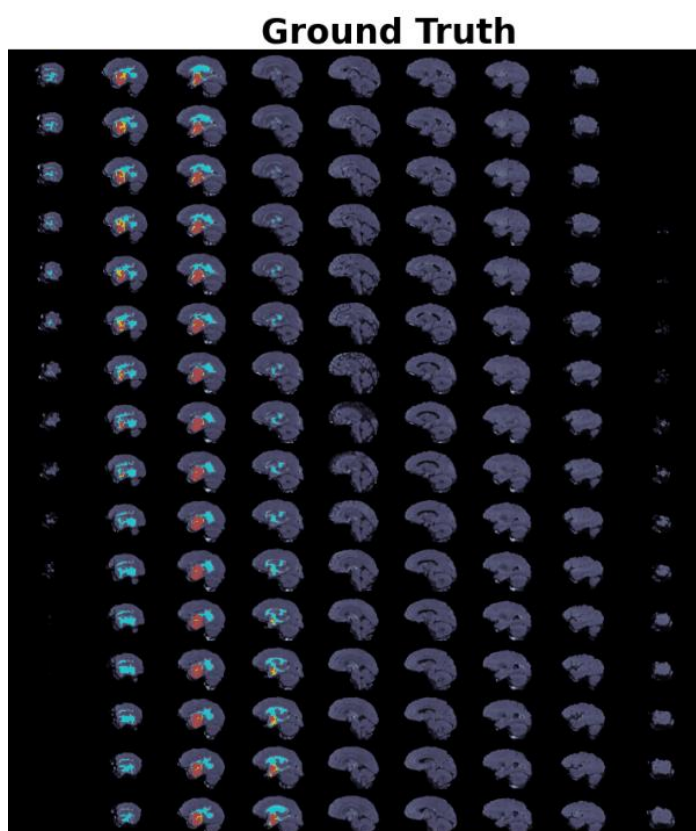


Figure 5. 13: Ground truth of a Patient's Brain tumor

In the figures 5.11 and 5.12 we can see the predictions on the 2D segmentations of the brain tumor compared to the ground truth of a single patient. In Fig.5.13 3D cross section of the segmented ground truth brain tumor of patient 184 and the Predicted 3D segmentation of the same patient is depicted in figure 5.14 using the test results.

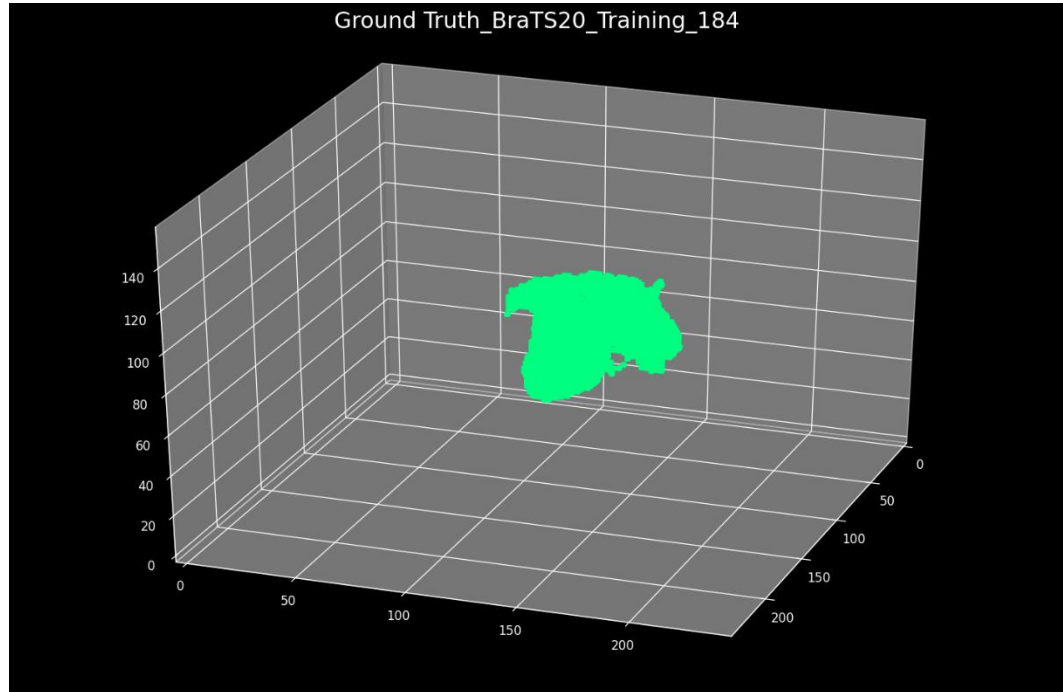


Figure 5. 4: 3D ground truth of a Brain tumor

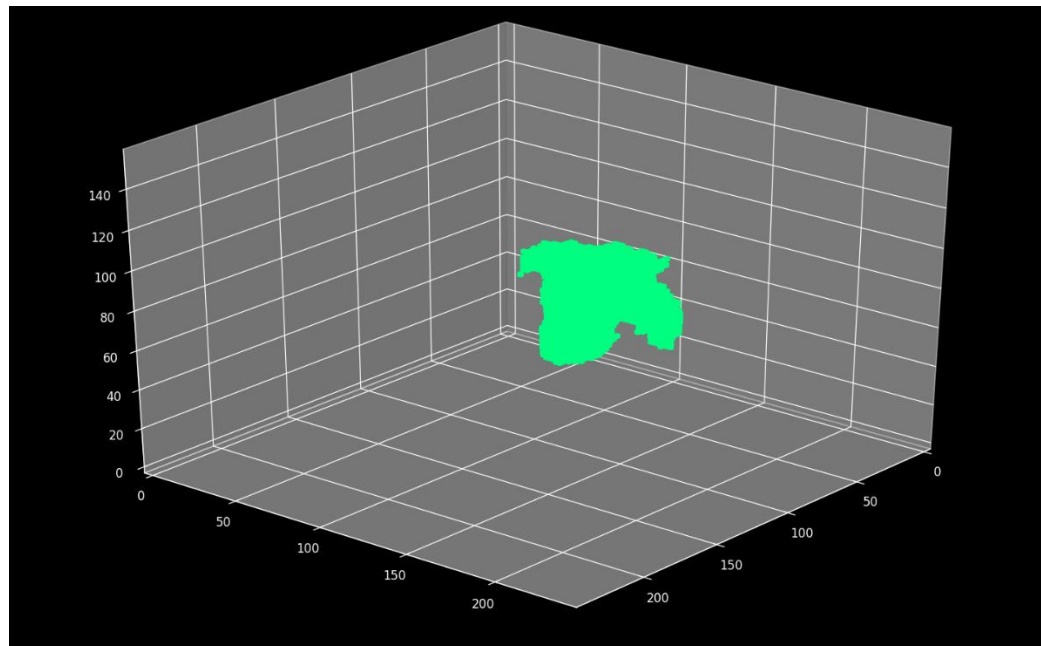


Figure 5. 15: 3D prediction of a Brain tumor

CHAPTER 06

CONCLUSIONS

In this work I present several Deep CNNs to address the task of automatic tumor segmentation from magnetic resonance images, with the objective of accelerating and improving radiotherapy planning and monitoring of patients with gliomas of varied grades. I started with a baseline architecture, U-Net, and gradually improved its performance and behavior of the architecture and adhering to its architecture adding a residual network (PT-Res-Net) which was pre trained and the cascaded system. I converted the multi-class segmentation problem to three cascaded binary segmentation problems and use three networks for segmentation. The network uses an anisotropic structure, which considers the balance among receptive field, model complexity and memory consumption. And an ensemble model with a majority rule method.

Results obtained from training these architectures in different datasets suggest that the addition of residual elements and the improved cascaded architecture boost the ensemble model for a better segmentation performance. However, an in-depth investigation of the intermediate representations learned by these networks is required in order to assess the changes in behavior attributed to these additions. Maybe, it's that interpretability is crucial in deep learning methods, and even more in clinical applications. The presented architectures demonstrate competitive performance compared to state-of-the-art methods when trained on the BraTS dataset. More robust version of these models was obtained via ensembling.

In the process of finding the best modality combination for a higher accuracy in brain tumor segmentation compared to all four modalities, I figured out the following point,

- Best modality combination was FLAIR, T1, and T2.
- Least impactful modality on the segmentation was T1-ce.
- Highest impactful modality on the segmentation was the FLAIR modality.

CHAPTER 07

FUTURE WORK

In this section, we discuss several possible future research directions.

7.1 More effective loss function for training deep CNNs

The current loss function for training deep CNNs is the cross-entropy loss function and its simple modification, one of the possible directions is to develop more effective loss function that can better focus on the tumor pixels and differentiate them from background pixels

7.2 Survival Prediction Feature Extraction

The current data is insufficient for the development of an accurate prediction of the survival of a patient in the dataset. So one possibility is to find more survival data from older BraTS challenge datasets as it spans from 2017 and improve a trained model for the new dataset with hopefully a higher number of patient data. And adhere to transfer learning techniques to obtain better accuracy.

7.3 Generative Adversarial Network for missing modality Reconstruction

Carry out experiments with the use of data augmentation, image reconstruction, image enhancement and missing value generation. Using image reconstruction techniques, analyze the percentage value of pixels needed for reconstructing the FLAIR and T2 modality. Data augmentation techniques will be utilized to evaluate the features needed to reconstruct the missing modalities. Developing a Generative Adversarial Network (GAN) to make a generative model for the reconstruction of the modalities will be utilized.

CHAPTER 08

REFERENCES

- [1] S. Asia and S. Asia, “Central nervous system,” *Pharm. Times*, vol. 78, no. 3, pp. 7–8, 2012.
- [2] “What Causes Brain Tumors? | Causes of Brain Cancer.” <https://www.cancer.org/cancer/brain-spinal-cord-tumors-adults/causes-risks-prevention/what-causes.html> (accessed Jul. 09, 2022).
- [3] A. Işin, C. Direkoğlu, and M. Şah, “Review of MRI-based Brain Tumor Image Segmentation Using Deep Learning Methods,” *Procedia Comput. Sci.*, vol. 102, pp. 317–324, 2016, doi: 10.1016/J.PROCS.2016.09.407/REVIEW_OF_MRI_BASED_BRAIN_TUMOR_IMAGE_SEGMENTATION_USING_DEEP_LEARNING_METHODS.PDF.
- [4] D. Lachinov, E. Vasiliev, and V. Turlapov, “Glioma segmentation with cascaded UNet,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11384 LNCS, pp. 189–198, 2019, doi: 10.1007/978-3-030-11726-9_17.
- [5] B. H. Menze *et al.*, “The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS),” *IEEE Trans. Med. Imaging*, vol. 34, no. 10, pp. 1993–2024, Oct. 2015, doi: 10.1109/TMI.2014.2377694.
- [6] Z. Akkus, A. Galimzianova, A. Hoogi, D. L. Rubin, and B. J. Erickson, “Deep Learning for Brain MRI Segmentation: State of the Art and Future Directions,” *J. Digit. Imaging*, vol. 30, no. 4, pp. 449–459, 2017, doi: 10.1007/s10278-017-9983-4.
- [7] “MRI Basics.” [https://case.edu/med/neurology/NR/MRI Basics.htm](https://case.edu/med/neurology/NR/MRI%20Basics.htm) (accessed Jul. 10, 2022).
- [8] S. Bauer, R. Wiest, L. P. Nolte, and M. Reyes, “A survey of MRI-based medical image analysis for brain tumor studies,” *Phys. Med. Biol.*, vol. 58, no. 13, pp. 1–44, 2013, doi: 10.1088/0031-9155/58/13/R97.
- [9] A. Esteva *et al.*, “Deep learning-enabled medical computer vision,” *npj Digit. Med.*, vol. 4, no. 1, pp. 1–9, 2021, doi: 10.1038/s41746-020-00376-2.

- [10] Y. Shen and M. Gao, “Brain Tumor Segmentation on MRI with Missing Modalities,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11492 LNCS, pp. 417–428, 2019, doi: 10.1007/978-3-030-20351-1_32.
- [11] A. Işin, C. Direkoğlu, and M. Şah, “Review of MRI-based Brain Tumor Image Segmentation Using Deep Learning Methods,” *Procedia Comput. Sci.*, vol. 102, no. August, pp. 317–324, 2016, doi: 10.1016/j.procs.2016.09.407.
- [12] M. Islam, V. S. Vibashan, V. J. M. Jose, N. Wijethilake, U. Utkarsh, and H. Ren, “Brain tumor segmentation and survival prediction using 3d attention unet,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11992 LNCS, pp. 262–272, 2020, doi: 10.1007/978-3-030-46640-4_25.
- [13] N. B. Bahadure, A. K. Ray, and H. P. Thethi, “Image Analysis for MRI Based Brain Tumor Detection and Feature Extraction Using Biologically Inspired BWT and SVM,” *Int. J. Biomed. Imaging*, vol. 2017, 2017, doi: 10.1155/2017/9749108.
- [14] B. Devkota, A. Alsadoon, P. W. C. Prasad, A. K. Singh, and A. Elchouemi, “Image Segmentation for Early Stage Brain Tumor Detection using Mathematical Morphological Reconstruction,” *Procedia Comput. Sci.*, vol. 125, no. 2009, pp. 115–123, 2018, doi: 10.1016/j.procs.2017.12.017.
- [15] F. Ning, D. Delhomme, Y. LeCun, F. Piano, L. Bottou, and P. E. Barbano, “Toward automatic phenotyping of developing embryos from videos,” *IEEE Trans. Image Process.*, vol. 14, no. 9, pp. 1360–1371, Sep. 2005, doi: 10.1109/TIP.2005.852470.
- [16] C. Farabet, C. Couprie, L. Najman, and Y. Lecun, “Learning hierarchical features for scene labeling,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1915–1929, 2013, doi: 10.1109/TPAMI.2012.231.
- [17] E. Shelhamer, J. Long, and T. Darrell, “Fully Convolutional Networks for Semantic Segmentation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Nov. 2014, doi: 10.48550/arxiv.1411.4038.
- [18] D. Ciregan, U. Meier, and J. Schmidhuber, “Multi-column deep neural networks for image classification,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 3642–3649, 2012, doi: 10.1109/CVPR.2012.6248110.

- [19] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, Sep. 2014, doi: 10.48550/arxiv.1409.1556.
- [20] W. Weng and X. Zhu, “INet: Convolutional Networks for Biomedical Image Segmentation,” *IEEE Access*, vol. 9, pp. 16591–16603, May 2021, doi: 10.1109/ACCESS.2021.3053408.
- [21] Y. Choi, Y. Kwon, H. Lee, B. J. Kim, M. C. Paik, and J. H. Won, “Ensemble of deep convolutional neural networks for prognosis of ischemic stroke,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10154 LNCS, pp. 231–243, 2016, doi: 10.1007/978-3-319-55524-9_22.
- [22] A. Casamitjana, S. Puch, A. Aduriz, and V. Vilaplana, “3D Convolutional Neural Networks for Brain Tumor Segmentation: A Comparison of Multi-resolution Architectures,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10154 LNCS, pp. 150–161, May 2017, doi: 10.48550/arxiv.1705.08236.
- [23] F. Yu and V. Koltun, “Multi-scale context aggregation by dilated convolutions,” Nov. 2016, doi: 10.48550/arxiv.1511.07122.
- [24] T. Pohlen, A. Hermans, M. Mathias, and B. Leibe, “Full-Resolution Residual Networks for Semantic Segmentation in Street Scenes,” *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-January, pp. 3309–3318, Nov. 2016, doi: 10.48550/arxiv.1611.08323.
- [25] N. K. Subbanna, D. Precup, D. L. Collins, and T. Arbel, “Hierarchical probabilistic Gabor and MRF segmentation of brain tumours in MRI volumes,” *Med. Image Comput. Comput. Assist. Interv.*, vol. 16, no. Pt 1, pp. 751–758, 2013, doi: 10.1007/978-3-642-40811-3_94.
- [26] “Proceedings of the MICCAI Challenge on Multimodal Brain Tumor Image Segmentation (BRATS) 2013 - Inria.” <https://hal.inria.fr/hal-00912934> (accessed Jul. 10, 2022).
- [27] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” Accessed: Jul. 10, 2022. [Online]. Available: <http://code.google.com/p/cuda-convnet/>.

- [28] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 770–778, 2016, doi: 10.1109/CVPR.2016.90.
- [29] “Segmentation of Brain Tumor Tissues with Convolutional Neural Networks - Microsoft Research.” <https://www.microsoft.com/en-us/research/publication/segmentation-of-brain-tumor-tissues-with-convolutional-neural-networks/> (accessed Jul. 10, 2022).
- [30] M. Noori, A. Bahri, and K. Mohammadi, “Attention-guided version of 2D UNet for automatic brain tumor segmentation,” *2019 9th Int. Conf. Comput. Knowl. Eng. ICCKE 2019*, pp. 269–275, 2019, doi: 10.1109/ICCKE48569.2019.8964956.
- [31] Y. Xue *et al.*, “A fully 3D multi-path convolutional neural network with feature fusion and feature weighting for automatic lesion identification in brain MRI images,” 2019, [Online]. Available: <http://arxiv.org/abs/1907.07807>.
- [32] M. Salvi *et al.*, “Integration of Deep Learning and Active Shape Models for More Accurate Prostate Segmentation in 3D MR Images,” *J. Imaging*, vol. 8, no. 5, p. 133, May 2022, doi: 10.3390/JIMAGING8050133.
- [33] P. K. Shukla, M. Zakariah, W. A. Hatamleh, H. Tarazi, and B. Tiwari, “AI-DRIVEN Novel Approach for Liver Cancer Screening and Prediction Using Cascaded Fully Convolutional Neural Network,” *J. Healthc. Eng.*, vol. 2022, 2022, doi: 10.1155/2022/4277436.
- [34] P. Ahmad, S. Qamar, L. Shen, and A. Saeed, “Context Aware 3D UNet for Brain Tumor Segmentation,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12658 LNCS, pp. 207–218, 2021, doi: 10.1007/978-3-030-72084-1_19.
- [35] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, “3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9901 LNCS, pp. 424–432, Jun. 2016, doi: 10.48550/arxiv.1606.06650.
- [36] S. Ioffe and C. Szegedy, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,” *32nd Int. Conf. Mach. Learn. ICML 2015*, vol. 1, pp. 448–456, Feb. 2015, doi: 10.48550/arxiv.1502.03167.

- [37] N. Srivastava, G. Hinton, A. Krizhevsky, and R. Salakhutdinov, “Dropout: A Simple Way to Prevent Neural Networks from Overfitting,” *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, 2014.
- [38] M. K. Panda, A. Sharma, V. Bajpai, B. N. Subudhi, V. Thangaraj, and V. Jakhetiya, “Encoder and decoder network with ResNet-50 and global average feature pooling for local change detection,” *Comput. Vis. Image Underst.*, vol. 222, p. 103501, Sep. 2022, doi: 10.1016/J.CVIU.2022.103501.
- [39] L. Sun, S. Zhang, H. Chen, and L. Luo, “Brain tumor segmentation and survival prediction using multimodal MRI scans with deep learning,” *Frontiers in Neuroscience*, vol. 13, no. JUL. 2019, doi: 10.3389/fnins.2019.00810.
- [40] D. P. Kingma and J. L. Ba, “Adam: A Method for Stochastic Optimization,” *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, Dec. 2014, doi: 10.48550/arxiv.1412.6980.

CHAPTER 09

APPENDIX

APPENDIX A:- Architecture Diagrams

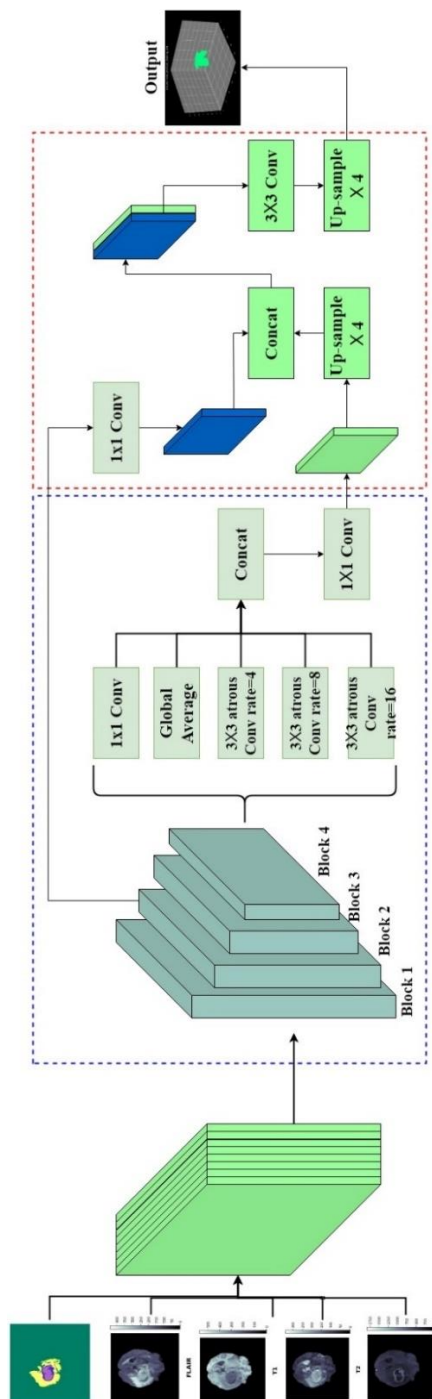


Figure 9. 1: Residual Network Architecture

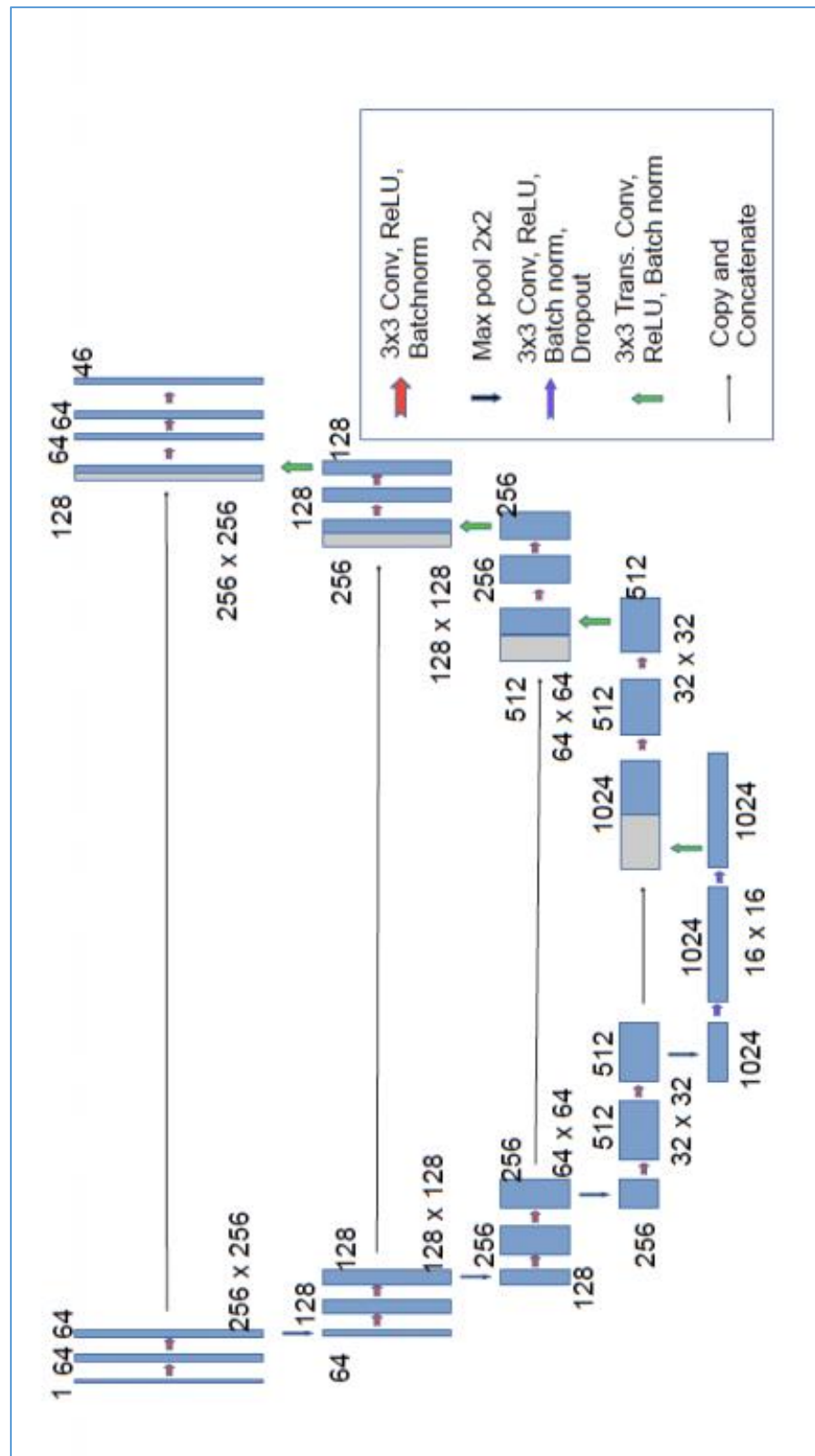


Figure 9. 2: U-Net Architecture

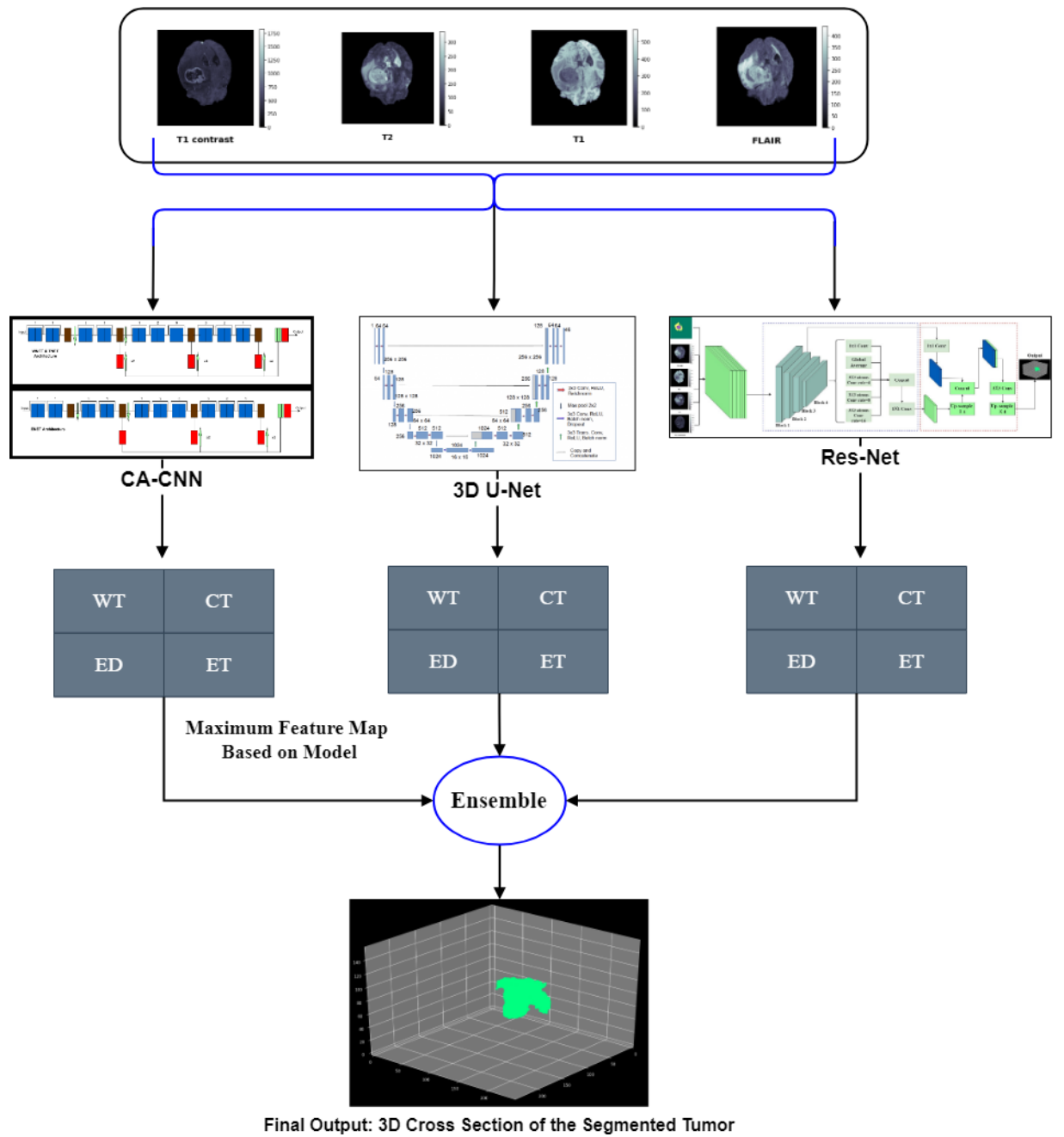


Figure 9. 4: Ensemble Architecture