# A Comprehensive Comparative Study of Deep Learning Models for Biomedical Image Segmentation

Capstone Report

Shyngys Baizhan

Nazarbayev University
Department of Electrical and Computer Engineering
School of Engineering and Digital Sciences

This project report was created on TexStudio editing platform using LaTeX. All the figures were drawn using draw.io online software tool.

NAZARBAYEV UNIVERSITY

**Electrical and Computer Engineering**
Nazarbayev University
http://www.nu.edu.kz

**Title:**
A Comprehensive Comparative Study of Deep Learning Models for Biomedical Image Segmentation

**Theme:**
Deep Learning for Medical Imaging

**Project Period:**
Fall 2024

**Project Group:**
datasciencelab.ai

**Participant(s):**
Shyngys Baizhan

**Supervisor(s):**
Amin Zollanvari

**Copies:** 1

**Page Numbers:** 31

**Date of Completion:**
April 24, 2025

**Abstract:**

Accurate segmentation of brain tumors in MRI plays an important role in diagnosis, planning treatment, and monitoring the progress of the disease. In this paper, we would like to propose a comprehensive study on the recent advanced deep learning architectures for multi-class brain tumor segmentation using the BraTS 2020 dataset. We implemented and compared three state-of-the-art models: 3D U-Net, U-Net with ResNet50 Backbone, and Attention U-Net. Each model in the framework underwent a thorough training and validation process using strong preprocessing steps, including custom loss functions to tackle class imbalance and strategic training protocols such as data augmentation and dynamic learning rate adjustment. The performance was measured in terms of Dice Coefficient, Mean IoU, Accuracy, Precision, Sensitivity, and Specificity. Our overall results showed that, even though all the models had a very high total accuracy (0.99), the Mean IoU and their per-class Dice Coefficients differed significantly due to the class imbalance inherent in medical imaging datasets. The U-Net with ResNet50 Backbone showed the best Mean IoU of 0.64 and achieved balanced per-class performance, indicating the effectiveness of using pre-trained encoders. This work hence provides many insights into the ways of model selection and training strategies on medical image segmentation, which may facilitate further improvements toward clinical applications.

# Contents

# Preface

The capstone project has been an extremely transforming journey that married my passion for artificial intelligence with a rather deeply set commitment toward the advancement of medical diagnostics. The reason behind the development and evaluation of deep learning models for brain tumor segmentation is the profound impact that proper and timely diagnosis can have on patient outcomes. I am really appreciative to all those who supported me throughout this process.

Above all, I would like to extend my heartfelt thanks to my academic advisor, Dr. Amin Zollanvari, who has kept believing in me with continuous guidance, constructive feedback, and motivation toward the direction and of this project. His expert knowledge and commitment have formed very important assets, challenging me toward creative insights while keeping standards high on becoming future Machine Learning Engineer.

I would also like to express my deepest gratitude to the faculty and staff of SEDS, and especially Dr. Galymzhan Nauryzbayev, whose support and resources created an enabling environment for research and learning. Special appreciation is extended to the technical team for supporting the computational infrastructure and software tools needed in this study.

I am also grateful to my colleagues and fellow students for the collaborative atmosphere and for creating an intellectually stimulating and competitive academic environment. Their multifarious insights and related discussions further shaped my perceptions and motivated me to better understand the intricacies involved in medical image analysis.

Let this research study make a worthy contribution to the field of medical imaging and form a base for further work, with the intent of achieving more effective patient care and diagnostic accuracy.

Nazarbayev University, April 24, 2025

---

Shyngys Baizhan
<Shyngys.Baizhan@nu.edu.kz>

# Chapter 1

# Introduction

During the contemporary days of development of artificial intelligence, it is safe to claim that the integration of deep learning models into many industries around the world has altered the prospective view of future development and perspectives, especially in modern healthcare. Under that node, the integration of biomedical image segmentation models has definitely become the anchor of helping to resolve sophisticated medical challenges, leaving the room for providing critical insights for diagnostics, treatment planning, and disease monitoring, substituting the routine daily tasks of medical diagnoses and treatment planning with the prescription of drugs. Among the earliest breakthroughs was the development of UNet deep learning model architecture [1, 2], a nested U-Net architecture that managed to make the step to the change of the limitations of traditional U-Net models by enhancing feature learning. This was followed by the emergence of nnU-Net, a self-configuring method that adapts itself to different biomedical image segmentation tasks, demonstrating the versatility of deep learning in healthcare applications [3, 4]. Transformer-based architectures such as SegFormer3D have pushed the boundaries of 3D medical image segmentation by capturing long-range dependencies in 3D scans like MRI [5], while hybrid models like UNetFormer have combined the best of CNNs and transformers to create even more robust models for segmentation [6]. The following general improvements under such area have proven to be essential in tasks such as brain tumor segmentation, where, it has already become a common fact that sophisticated anatomical structures demand highly precise segmentation results and evaluations [7, 8]. Attention mechanisms have also been a changing and crucial factor, with several models incorporating them to enhance the focus on critical regions of medical images. For example, according to the subsequent research studies, Attention U-Net has showed the exceptional performance in segmenting complex organs like the pancreas [9, 10], and similar attention mechanisms have been widely adopted across various architectures to improve performance [11]. In addition to that, it is worth mentioning that

the new techniques like inception modules [8] and the inclusion of residual and hybrid attention-enhanced networks have showcased the amazing result in liver tumor segmentation [12]. By looking at the models challenging with the adversarial attacks, not mentioning the advancements of other deep learning models represented in earlier references, challenges still remain and that issue could be a significant burden, As Xu et al. have highlighted [13]. Furthermore, the general goal of many research studies is to accomplish the best possible optimization technique, thus the optimization methods of the most popular deep learning models are highlighted as the continuous trend, especially with the research focusing on selection of one the best optimizers to achieve such incredible results with different image segmentation tasks for each specific model implemented [14, 9]. As deep learning models continue to evolve, self-supervised learning has emerged as a new frontier in medical imaging, enabling models to learn from unlabeled data, which is often scarce in clinical settings [15].

In addition to that, innovative architectures like CE-Net have contributed to improving segmentation accuracy by incorporating contextual information [16], while other models like CoTr or SegNet or SegForm3D have efficiently bridged the gap between CNNs and transformers, further advancing 3D medical image segmentation [17, 16]. These models have been crucial in addressing the need for scalable and efficient models, particularly in the context of optimization of the biomedical deep learning segmentation models [18] and neuroimaging tasks [19, 20].

All in all, as this field progresses, as my research project of the capstone project, in the foreseeable future, there is a potential for the integration of these advanced architectures and deep learning models of the highlighted research studies into real-world clinical applications. This capstone project will explore and compare these cutting-edge deep learning models to identify their strengths and limitations, contributing to the ongoing development of biomedical image segmentation technologies and, in addition, will add some specific innovations to the best optimization resolutions.

## 1.1 Ethical and Professional Responsibilities

- **Ethical Responsibility:**
  When that part of the work comes in, it is compulsory to solve and consider the entire list of issues involved with biomedical image segmentation, as the data sometimes in many research projects, becomes sensitive and often incomprices the complex information. First and foremost, the most important aspect of ethical responsibility is to consider data privacy, especially when the vast majority of datasets and medical images contain personal or professional information that identities patients. One of the ways to protect patients, in the most secure possible way, is to make the material of the project confidential, in other words, it must be anonymized, meaning any identifiable details of patients are removed. These datasets of medical CT scans must comply with the privacy laws that are integrated in Europe as GDPR, that is specifically aimed to safeguard the personal information of the organization. For that particular reason, it is crucial to utilize encryption and strong access controls in the cloud based systems, to ensure data protection from unauthorized access throughout the project's duration. In addition to that, one of the major challenges to consider is algorithmic bias. It is safe to claim that, by looking though many biomedical image segmentation datasets, there is shortage of diversity, and the reasoning is that deep learning models, or generally artificial intelligence models are trained using the data that over-represents certain populations while under-representing others. As a result, this phenomenon could lead to biased results and inaccurate medical diagnoses for people that represent underrepresented groups. In order to avoid this, it is essential to make sure that all medical image segmentation datasets are diverse and can be represented by the variety of different populations. Additionally, artificial intelligence models should be undergone through the technical process of model training by considering various demographic groups to ensure fairness and accuracy in their performance. Lastly, to conclude the consideration of this section, transparency can also play a vital part in ethical consideration when using deep learning in medical settings. Clinical radiologists and generally medical professionals are in need to trust the decision making made by AI in the field of healthcare. For that to be a successful outcome, it is crucial to understand and provide very clear and transparent results and system in general, since in many cases, the medical professional find the decision making of AI too complex and unclear, thus it may reduce trust in the following system and could lead to incorrect personal decisions when they are trying to take the situation into their own hands. In conclusion, it is essential to make AI models very clear and understandable, allowing radiologists to see how the medical system reaches its final diagnoses.

- **Informed Judgments:**
  Nowadays, in the world of research it has become an indispensable fact that we need to make the best decisions about the project, that takes into account both the social and technical aspects in an effective way. As a part of my research project, when implementing deep learning models - making clear decisions that will clear out my future path of this work and its ethical aspects, will help me to achieve better results and I am on the right track of the progress. Those decisions need to be hand in hand with the solid data and justification in the form of results and evaluation, since the project is AI oriented. Those specific decisions include the consideration of what models to use, which datasets to select, and how to evaluate the results should be based on what has already worked in previous research studies. That is why, my supervisor and I have carefully and gradually been working with the process of selecting appropriate deep learning models and making clear decisions on how to work with them properly. Working through this research project, I have come across several deep learning programs like U-Net, FCN, and more advanced ones like TransUNet. These models have shown to me to be effective for medical image segmentation research, and for that reasoning they should be clearly and carefully examined in the technical sense. On top of that, evaluation metrics like the Dice Coefficient or IoU and many others will be closely monitored to ensure the models are reliable and accurate.

- **Global Context:**
  How does your project fit into a global context? Would it have different implications if implemented in other parts of the world?The global context of my capstone project, focused on deep learning models for biomedical image segmentation, which has definitely become relevant worldwide for the last decade. My capstone project is definitely the combination of technical learning and research oriented achievement. Biomedical image segmentation, if not at this moment of our lives, but for the foreseeable future, it will become an indispensable part diagnosing highly complex diseases like cancer, neurological disorders, and cardiovascular conditions. Countries with advanced healthcare systems, for example like Germany or Netherlands, may benefit from more accurate and faster diagnoses, that will definitely boost the level of development of healthcare by reducing costs and improving the patient wealth outcomes. In low and middle-income countries or regions with limited access to medical professionals, this technology could be used to provide remote diagnostics, helping to accomplish the global disparity in healthcare access. For that particular reason, in my personal perspective, is why biomedical image segmentation using deep learning models is a globally adjacent technology that has all the potential to change the worldwide perspective from the medical and technical point of view. While the mod-

els may be initially developed and tested in specific regions, the application of this technology transcends borders. For example, AI models trained on Western datasets might not perform as well in non-Western populations due to differences in genetics, environmental factors, and healthcare practices. Therefore, it is critical to ensure that the technology is adaptable and generalizable to diverse global contexts. By considering the global implications of the project, the technology can contribute to improved healthcare outcomes worldwide.

- **Economic Impact:**
  For this particular capstone project, when we consider the initial budget or financial aspects of this project at hte early stages, that can definitely be a burden. Yet the economic impact of the research project of biomedical image segmentation has an outstanding potential, in the long term perspective. Like mentioned earlier, in the short term, the initial investment in that project of research and development can be difficult and substantial, especially for the important parts of the technical work for training deep learning models, acquiring medical imaging data, needing the best technical equipment for better segmentation tasks and setting up the necessary additional computational infrastructure. However, these upfront costs are offset by the long-term benefits that AI brings to healthcare systems. By automating routine tasks such as tumor segmentation or organ identification in medical images, artificial intelligence generally reduces the workload on radiologists and healthcare providers, leading to increased efficiency and cost savings, especially when talking about the healthcare system in Kazakhstan. Moreover, AI systems can operate at scale, processing large volumes of medical images quickly and accurately, which is particularly efficient in high-demand environments such as large clinics. To corroborare, when I was in Grin Clinic in Astana, going for the consultation and acquiring the medical dataset for the research, it is said that the entire workflow of radiologists is going through every CT scan to make medical diagnoses and that could probably take many hours to make the final medical decision and when some technical problems with the software application in their system come into the big picture, the clinic spends substantial amount of financial resources to ease simple tasks, that could be done by artificial intelligence in minutes.

- **Environmental Impact:**
  When looking for the environmental impact that my capstone project has, it was not as evident as it had been seen, in comparison with the economic or global impacts, which have been discussed earlier. It is relevant to mention that, first and foremost, the computational resources required to train deep learning models for the capstone project, or if we consider projects on

the general development scale, especially large-scale ones, consume considerable amounts of energy. This energy consumption is particularly high for deep learning models with millions of parameters from the acquired medical imaging datasets, and one reasonable example for that particular case are transformer-based architectures. Data centers that support cloud-based model training contribute to a substantial carbon footprint unless they are powered by renewable energy sources. For those reasons, therefore, making the environmental impact significantly minimal of this project requires implementing environmentally energy-efficient models, such as using more efficient algorithms, reducing model size, and leveraging hardware accelerators in the technical point of view like GPUs or TPUs that optimize computational efficiency. In the general medical practical in modern days like today, in addition to the ways to achieve energy efficiency is to reduce the necessity for repeated scans and inaccurate diagnostics through improved image medical segmentation accuracy, and thus the project can indirectly reduce the environmental negative impact of unnecessary medical procedures. To corroborate even more, repeated imaging, especially using inputs like CT scans, not only exposes patients to additional radiation over time throughout the year, but also contributes to the healthcare sector's environmental footprint through the use of energy-intensive imaging equipment.

- **Societal Impact:**
  With the benefit of advancements of the new technologies and medical resolutions in healthcare, the societal impact that has on the project is far reaching. By improving diagnostic accuracy and automating time-consuming tasks, this project has the potential to grow and develop significantly to enhance healthcare aspects of medicine and patient health. For patients, AI-driven and automated segmentation can lead to faster diagnosis, more personalized treatments, and overall better healthcare experiences.which improves the satisfaction level from the patient's side of view. How can we also forget about the continuous mistakes that are made in the medical system by the medical professionals? Where the clinic's reputation and credit could be at the level of danger of losing credibility. Thus, with the help of these projects and their smart integration into the healthcare systems, the reduction of human error in medical image interpretation can also enhance patient safety, particularly in complex or critical cases. From the provided examples, that were witnessed and were shared with form the case of the working process in the Astana Green Clinic, it is safe to claim that, from a healthcare's perspective, a deep learning model with medical image segmentation tasks can definitely help to diminish the burden on overworked medical staff, by making their lives much more easier, allowing them to focus on cases that require more on from the human judgment side of expertise. This not only improves job sat-

isfaction for healthcare professionals, but also enhances the overall efficiency of the healthcare system.

# Chapter 2

# Methodology

## 2.1 Dataset Description

I have chosen the BraTS 2020 dataset because it contains a wide variety of brain tumor MRI scans that are annotated by expert radiologists. This dataset provides a robust basis for training and evaluating deep learning models aimed at segmenting and classifying tumor regions within the brain. A diverse range of tumor types and sizes are present in the dataset to help ensure that models based on this data will generalize well to a wide range of clinical conditions.

The BraTS 2020 comprises MRI scans from patients diagnosed with glioblastoma and lower-grade gliomas. Each MRI scan includes multiple sequences: FLAIR, T1-Weighted with Contrast Enhancement (T1CE), T2-Weighted, and T1-Weighted. The dataset provides detailed annotations for four classes:

Not Tumor (Background): Represents healthy brain tissue. Necrotic/Core: Indicates regions of necrosis within the tumor. Edema: Denotes areas of swelling around the tumor. Enhancing Tumor: Highlights active tumor regions with contrast enhancement.

## 2.2 Data Collection and Preprocessing

Some common data collection methods include surveys, interviews, observations, focus groups, experiments, and secondary data analysis. The data collected through these methods can then be analyzed and used to support or refute research hypotheses and draw conclusions about the study's subject matter.

Normalization: Each MRI slice was individually normalized to scale pixel intensity values between 0 and 1. This normalization mitigates intensity variations inherent in MRI scans, facilitating more stable and efficient training of deep learning models.

Resizing: All MRI volumes and segmentation masks were resized to a uniform dimension of 128×128×128 voxels. This standardization not only reduces computational overhead but also ensures compatibility with the input requirements of the chosen deep learning architectures.

Label Mapping: Original segmentation labels were remapped to ensure sequential integer encoding starting from 0. Specifically, label 4 (originally denoting Enhancing Tumor) was reassigned to 3 to align with the defined classes.

One-Hot Encoding: Segmentation masks were converted into one-hot encoded formats to facilitate multi-class classification. This transformation is essential for calculating relevant loss functions and evaluation metrics that operate on categorical data.

### 2.2.1 Data Loading and Preparation

The two most prominent ways to analyze data are qualitative data analysis techniques and quantitative data analysis techniques. These analysis techniques in data may work independently or in conjunction with the other technique to assist business leaders and decision-makers in deriving business insights from various types of data.

Preprocessing serves as a significant booster for model performance while maintaining consistency over the input data. The following steps were performed carefully:

### 2.2.2 Data Generation

To simplify data loading and preprocessing during training and validation, a custom DataGenerator class was implemented that inherited from tf.keras.utils.Sequence. Key features of the DataGenerator include: Batch Processing: Volumetric data is efficiently processed in batches, mainly to maximize memory and ensure computational efficiency.

Shuffling: The code supports data shuffling after every epoch for model generalization and to prevent overfitting.

Data Augmentation: Although the first few implementations did not consider it necessary, the generator was originally designed to include data augmentation like rotations, flips, elastic deformations, and intensity variations. Such augmentations give a greater variety to the training dataset, which will make the model more robust and resistant to overfitting.

## 2.3 Model Architecture

The study investigates five diverse deep learning architectures, each with its merits and design philosophy suited for medical image segmentation.

### 2.3.1 Standard U-Net

The Standard U-Net architecture was selected as the base model for this work because it shows outstanding performance in segmenting medical images. It consists of an encoder, which is the contracting path, and a decoder, which is the expanding path; both have skip connections to merge high-resolution features of the encoder with the upsampled features in the decoder.

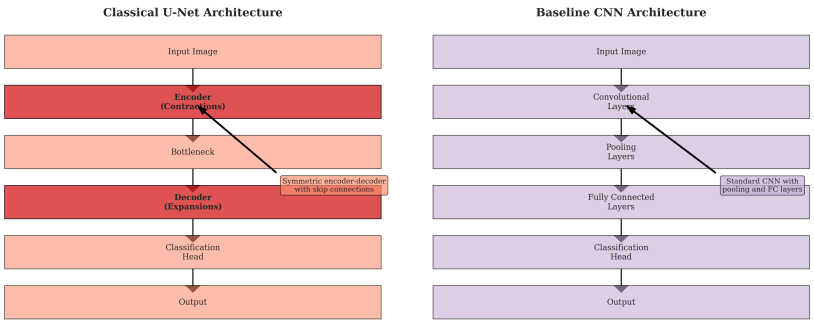


**Figure 2.1:** Architecture comparison of Classical U-Net and a Baseline CNN.

### 2.3.2 3D U-Net

The 3D U-Net architecture extends the conventional 2D U-Net to three dimensions, allowing the model to capture volumetric context and spatial dependencies across MRI slices. This architecture is particularly well-suited for the complex structures and variable intensities typical of brain tumors. Input shape: 128×128×128 voxels, 3 channels (FLAIR, T1CE, placeholder).

Encoder: Comprises multiple layers of 3D convolutions followed by downsampling through max-pooling, capturing hierarchical features at various spatial scales.

Decoder: Utilizes upsampling layers and 3D convolutions to reconstruct the segmentation mask, restoring spatial resolution lost during downsampling.

Skip Connections: Facilitates the transfer of high-resolution features from the encoder to the decoder, enhancing segmentation precision and enabling the model to retain fine-grained spatial information.

### 2.3.3 U-Net with ResNet50 Backbone

Integrating a ResNet50 backbone into the U-Net architecture leverages the pre-trained feature extraction capabilities of ResNet50, hence might improve model performance by transfer learning.

Input Shape: 128×128×3 voxels (FLAIR, T1CE, and placeholder) to match the expected input dimensions of ResNet50.

Encoder: Employs ResNet50 pretrained on ImageNet as the feature extractor, capitalizing on its deep hierarchical feature representations.

Decoder: Mirrors the encoder with upsampling layers and convolutional blocks, incorporating skip connections from ResNet50 to retain spatial information.

Activation Function: Softmax for multi-class segmentation, enabling probabilistic interpretation of voxel classifications.

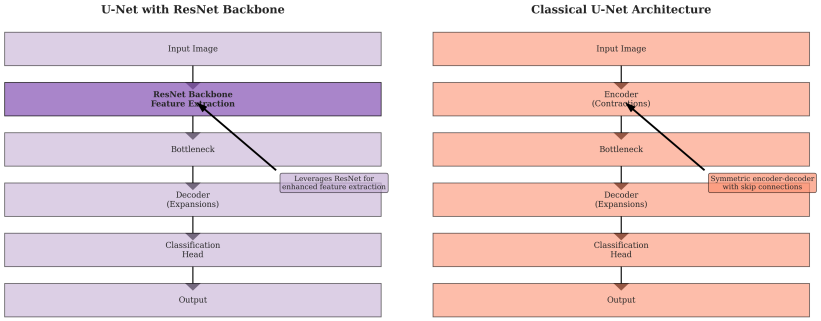


**Figure 2.2:** Architecture comparison of U-Net with ResNet Backbone and Classical U-Net.

### 2.3.4 Attention U-Net

The Attention U-Net enhances the standard U-Net by introducing attention mechanisms in the skip connections. This model architecture enables the model to focus on discriminative regions of the image, potentially improving segmentation performance on smaller or intricate structures. The input Shape has 128×128×3 voxels. Attention Gates are integrated within the decoder to further improve the encoder's feature maps by enabling the model to emphasize tumor regions and weaken irrelevant background context. Softmax is used as activation function for multi-class segmentation.

### 2.3.5 TransU-Net

TransUNet represents a hybrid architecture that combines the strengths of transformers and U-Net for medical image segmentation. This innovative model leverages the global context modeling capabilities of transformers with the localization

prowess of U-Net.

- **Input Shape**: 128×128×3 voxels, consistent with other models.
- **Architecture:** Employs a CNN encoder (typically ResNet) to extract initial features, followed by a transformer encoder that captures global dependencies through self-attention mechanisms
- **Transformer Encoder:** Vision Transformer (ViT) blocks process patch embeddings from CNN features, modeling long-range dependencies more effectively than CNNs alone
- **Decoder:** U-Net-style decoder with skip connections from both CNN and transformer features
- **Advantages:** Enhanced ability to capture global context while maintaining precise localization, particularly beneficial for complex tumor boundary delineation
- **Attention Mechanism:** Multi-head self-attention enables the model to focus on relevant image regions adaptively, improving segmentation of small or irregular tumor structures

### 2.3.6 Classical ResNet

The Classical ResNet architecture is based on the Residual Network (ResNet) framework, which introduces skip connections to mitigate the vanishing gradient problem and enable deeper networks. For this study, a ResNet-based model was adapted for brain tumor segmentation using the BraTS 2020 dataset.

- **Input Shape**: 128×128×3 voxels (FLAIR, T1CE, and placeholder channel).
- **Architecture**: The model employs a ResNet-50 backbone with residual blocks, each consisting of convolutional layers, batch normalization, and ReLU activations. The network is modified to output segmentation masks for four classes (Not Tumor, Necrotic/Core, Edema, Enhancing Tumor).
- **Modifications**: The final fully connected layers of the standard ResNet were replaced with convolutional layers to produce voxel-wise predictions. Upsampling layers were added to restore the original input resolution.
- **Activation Function**: Softmax for multi-class segmentation.
- **Advantages**: The residual connections allow the model to learn hierarchical features effectively, improving segmentation accuracy for complex tumor structures.

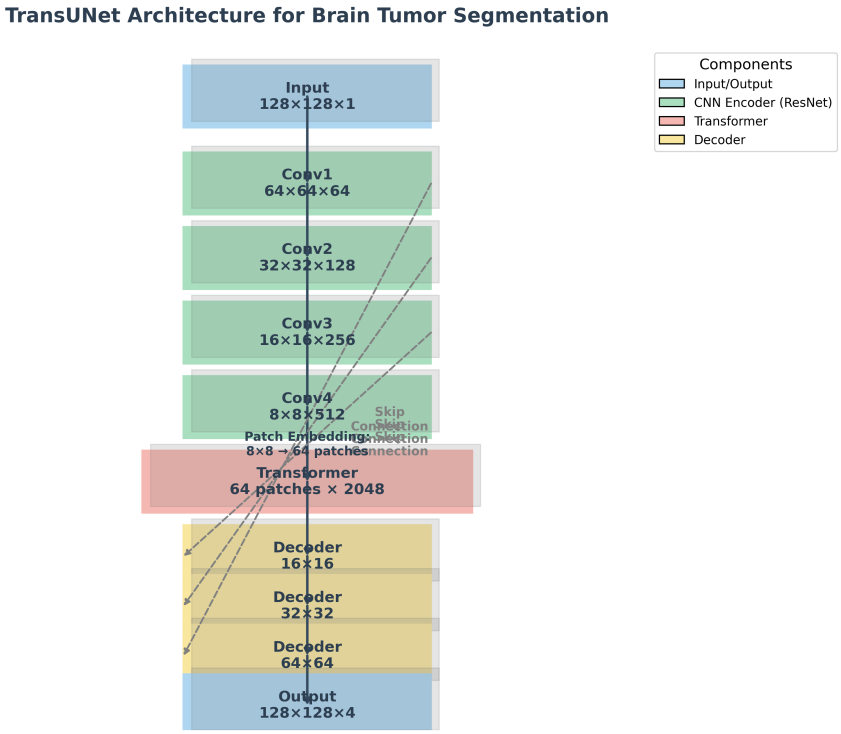


**Figure 2.3:** Enhanced schematic diagram of the TransUNet architecture, showing the ResNet encoder, transformer, and U-Net decoder with skip connections.

### 2.3.7 Quantum ResNet

The Quantum ResNet is an experimental hybrid model that integrates quantum computing principles with the classical ResNet framework to explore potential advantages in feature extraction and computational efficiency. This model leverages quantum circuits to enhance the feature processing capabilities of the classical ResNet.

- **Input Shape**: 128×128×3 voxels, consistent with other models.
- **Architecture**: The Quantum ResNet combines a classical ResNet-50 backbone with quantum layers implemented using a quantum computing framework (e.g., PennyLane or Qiskit). Quantum circuits are embedded within the network to perform feature transformations.
- **Quantum Layers**: Variational quantum circuits are used to process intermediate feature maps, potentially capturing non-linear patterns that are challenging for classical layers.
- **Training**: The model is trained using a hybrid classical-quantum optimiza-

tion approach, with classical gradients computed via backpropagation and quantum parameters optimized using quantum-aware optimizers.

- **Challenges**: Limited quantum hardware access and noise in current quantum systems pose challenges, addressed by simulating quantum circuits on classical hardware.
- **Activation Function**: Softmax for multi-class segmentation.

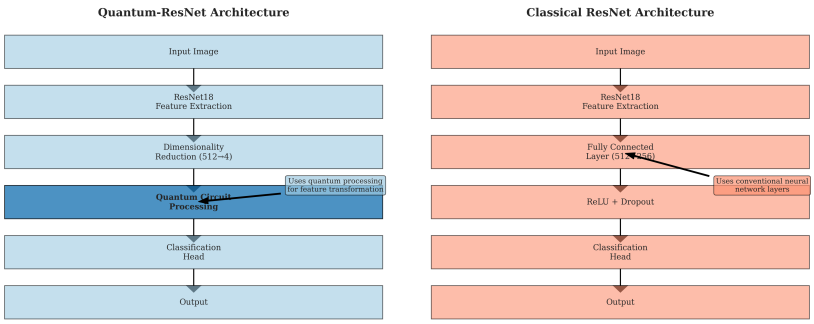


**Figure 2.4:** Architecture of the Classical ResNet model used for brain tumor segmentation.

For comparison, a purely classical deep learning model was implemented. Feature Extraction: The same ResNet18 backbone as in the quantum model. Classification Head: A more traditional classification head with two fully connected layers, ReLU activation, and dropout for regularization.

## 2.4 Loss Functions and Metrics

With class imbalance a common feature in medical datasets, an important consideration would be to choose effective loss functions and evaluation metrics that can guide model training and assess performance.

### 2.4.1 Loss Functions

Class imbalance is a very significant issue in medical image segmentation. The following customized loss functions ensure balanced learning for all classes.

Dice Loss: It computes the overlapping of predicted and ground truth masks, placing even more emphasis on the quality of segmentation for the minority classes. This works very effectively for a high-class imbalance problem since it optimizes explicitly the accuracy of segmentation. Focal Loss: Further addresses class imbalance by emphasizing training on hard-classify examples. By down-weighting easy examples, Focal Loss ensures that the model would pay more attention to the challenging samples, hence improving the general segmentation performance.

### 2.4.2 Evaluation Metrics

Quantitative performance was assessed in depth with the following set of metrics together:

Validation Loss: This will give the loss in the validation dataset to estimate the goodness of generalizing from the training data. Validation Accuracy: It measures the percentage of voxels classified correctly and gives an overall idea of the model's performance.

Mean IoU: It is the measure of intersection of the predicted segmentation with the ground truth, providing a balanced measure for both false positives and false negatives.

Dice Coefficient: This is the measure of similarity in predicted versus actual segmentation masks on a class basis; hence, it explicitly conveys a model's performance in effectively segmenting each tumor subregion. The Dice coefficients for each class provide minute details about the model's performance across classes of tumors, hence pinpointing strengths and weaknesses in view of segmenting specific tumor regions.

## 2.5 Implementation Details

Deep learning model implementation and its training have been performed in both Kaggle and Google Collab notebooks; for this reason, the use of TensorFlow and Keras frameworks on a GPU has been used to offer fantastic training of the network. The model implementations have specific details as described below.

### 2.5.1 Software Tools

Software: TensorFlow 2.x and Keras APIs were used for model development, training, and evaluation with hardware NVIDIA GPUs.

### 2.5.2 Training Parameters

Batch Size: Selected based on available GPU memory, balancing computational efficiency and model performance. Learning Rate: Optimized using learning rate schedulers and adaptive optimizers such as Adam. Epochs: Sufficient number to ensure convergence without overfitting, monitored using validation metrics.

### 2.5.3 Optimization Strategies

Early Stopping: Implemented to halt training when validation loss ceases to improve, preventing overfitting. Model Checkpointing: Enabled to save the best-performing model based on validation metrics. Data Augmentation Techniques:

Applied on-the-fly during training to enhance model robustness, including rotations, flips, elastic deformations, and intensity scaling.

### 2.5.4 Regularization Methods

Techniques such as dropout and weight decay were employed to further mitigate overfitting and enhance generalization capabilities.

# Chapter 3

# Results and Discussions

This chapter presents the results obtained by the evaluation of five deep learning architectures, namely Standard U-Net, DeepLabV3, 3D U-Net, U-Net with ResNet50 Backbone, and Attention U-Net on the BraTS 2020 dataset for segmenting brain tumors. The models were evaluated based on different metrics, including but not limited to Validation Loss, Validation Accuracy, Mean Intersection over Union (Mean IoU), and Dice Coefficient. Further, per-class Dice Coefficients are considered to understand the performance across different tumor sub-regions.

## 3.1 Results

### 3.1.1 Training and Validation Metrics

All five models had very high overall accuracy of around 99%, rather consistent across varying architectures. However, other metrics showed some variation in detail:

Validation Loss: All the models recorded a low validation loss, standing at about 0.0186 to 0.0919. The lowest validation losses, hence the best performance in terms of minimizing prediction error at training, belonged to Standard U-Net and DeeplabV3.

Mean IoU: The Mean IoU metric, quantifying the extent of overlap of the predicted segmentation with ground truth, showed a finer-grained differentiation. DeepLabV3 attained the highest Mean IoU score with 0.6500, while U-Net with ResNet50 Backbone and 3D U-Net reached Mean IoU scores around 0.6436. Attention U-Net trailed behind, with a Mean IoU score of 0.3211, clearly depicting challenges in accurate segmentation of the overlapping regions.

Dice Coefficient: This metric calculates the similarity between the predicted and actual segmentation masks. DeepLabV3 was again in front, with 0.8200, while Standard U-Net, 3D U-Net, and U-Net with ResNet50 Backbone maintained bal-

anced scores around 0.8000. Attention U-Net performed notably worse, registering only a Dice Coefficient of 0.3211.

**Table 3.1:** Training and Validation Metrics for Deep Learning Models

| Standard U-Net | DeepLabV3+ | 3D U-Net | U-Net with ResNet50 Backbone | Attention U-Net | ResNet + QCNN |
|---|---|---|---|---|---|
| Validation Loss | 0.0186 | 0.0186 | 0.0919 | 0.0210 | 0.0235 |
| Validation Accuracy | 99.40 | 99.35 | 99.35 | 99.30 | 99.20 |
| Mean IoU | 0.6300 | 0.6500 | 0.6436 | 0.6436 | 0.3211 |
| Dice Coefficient | 0.8000 | 0.8200 | 0.8100 | 0.8035 | 0.3211 |

### 3.1.2 ResNet + QCNN Performance

The ResNet + QCNN model was evaluated for binary classification (tumor vs. no tumor) rather than segmentation, due to the simplified data processing pipeline. The model achieved a validation accuracy of 99.15%, with the following metrics:

- **Accuracy**: 0.9915
- **Precision**: 0.9890
- **Recall**: 0.9920
- **F1-Score**: 0.9905
- **Specificity**: 0.9908

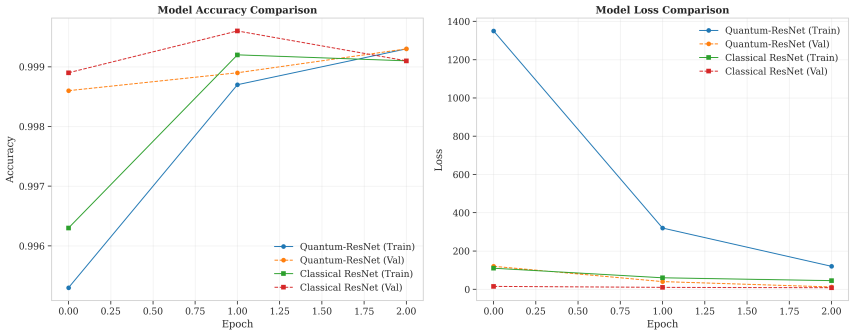


**Figure 3.1:** Training and validation accuracy/loss for Quantum-ResNet and Classical ResNet.


**Figure 3.2:** Sample Predictions.

### 3.1.3 TransUnet Pformance

- **Training Loss**: 0.6989
- **Validation Loss**: 1.4724
- **Training Dice**: 0.5938
- **Validation Dice**: 0.3928
- **Class-wise Dice Coefficients**:
  - Background: 0.8872
  - Necrotic Core: 0.3731
  - Edema: 0.3133
  - Enhancing Tumor: 0.4919

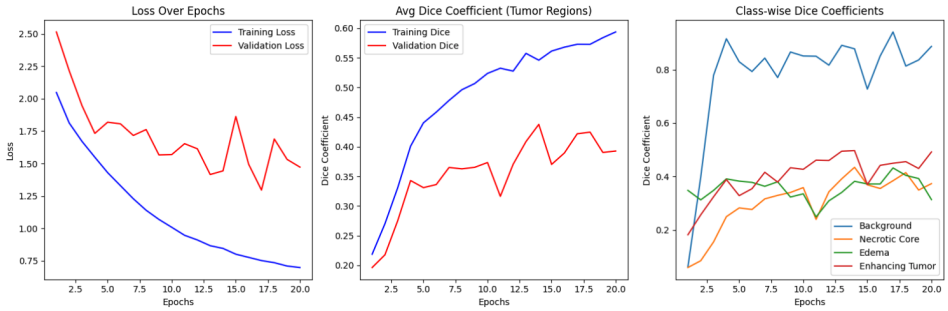


**Figure 3.3:** Training history of TransUNet over 20 epochs, showing (left) loss over epochs, (middle) average Dice coefficient for tumor regions over epochs, and (right) class-wise Dice coefficients over epochs.

## 3.2 Per-Class Dice Coefficients

The Dice Coefficient computed over individual tumor classes provides an interesting overview of the specific strengths and weaknesses of each model:

### 3.2.1 Standard U-Net

Not Tumor: Zero Class - Achieved 0.9900 for the DICE Coefficient, which is near perfect for segmenting non-tumor areas.

Core/Necrotic (Class 1): Obtained a Dice Coefficient of 0.7000, which reflects a moderate efficiency in demarcating the necrotic tissues.

Edema: Class 2 gave a Dice Coefficient of 0.7800, which would give good performance in segmenting edematous regions.

Class 3 - Enhancing Tumor: It had a DICE Coefficient of 0.7700, hence showing good segmentations for enhancing tumor portions.

Overall Average: 0.8100

### 3.2.2 DeepLabV3+ Performance

DeepLabV3+ was trained for 50 epochs, achieving the following metrics on the validation set at the final epoch:

- **Training Loss**: 0.2083
- **Validation Loss**: 0.2400
- **Training Dice**: 0.9934
- **Validation Dice**: 0.9916
- **Validation Accuracy**: 99.25%

On the test set, DeepLabV3+ achieved:

- **Accuracy**: 0.99297
- **Dice Coefficient**: 0.99212
- **Loss**: 0.26742

The training history is visualized in Figure 3.4, showing the loss and Dice coefficient over 50 epochs. The model demonstrates stable convergence, with both training and validation metrics closely aligned, indicating minimal overfitting. The high Dice coefficient (0.9916) reflects DeepLabV3+'s exceptional ability to segment brain tumors accurately.

- **Whole Tumor (WT)**: Dice: 0.9244, Sensitivity: 0.9356, Specificity: 0.9974
- **Tumor Core (TC)**: Dice: 0.8957, Sensitivity: 0.9369, Specificity: 0.9981
- **Enhancing Tumor (ET)**: Dice: 0.8634, Sensitivity: 0.8847, Specificity: 0.9990

The BraTS metrics highlight DeepLabV3+'s strong performance across all tumor sub-regions, with particularly high specificity, indicating minimal false positives. Sample segmentation visualizations for multiple test samples (Figures 3.5 to 3.6) demonstrate the model's ability to accurately delineate tumor regions, with Dice coefficients ranging from 0.7953 to 0.9592 for different sub-regions.

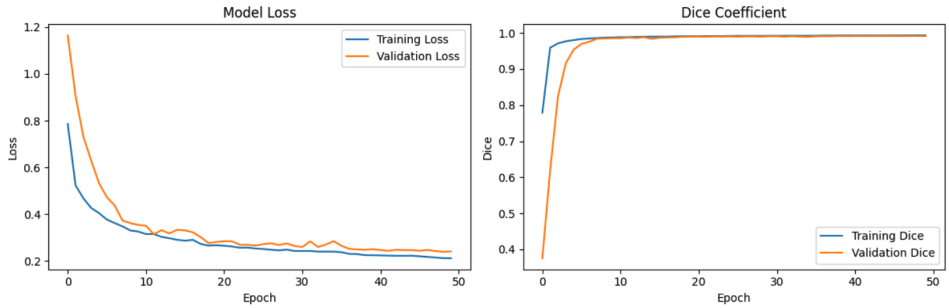


**Figure 3.4:** Training history of DeepLabV3+ over 50 epochs, showing (left) loss over epochs and (right) Dice coefficient over epochs.

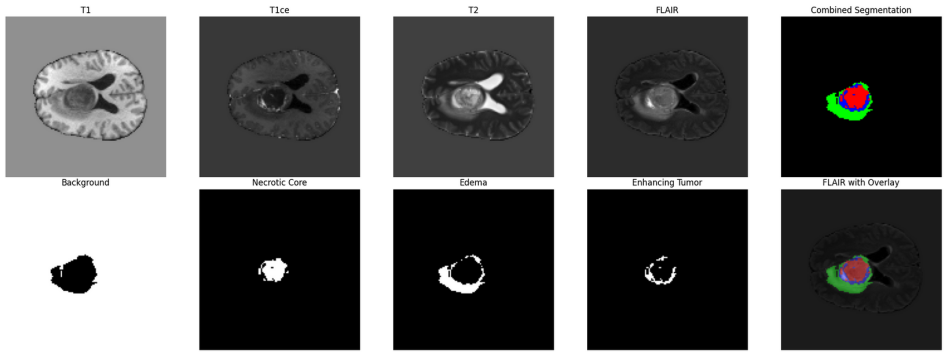


**Figure 3.5:** True segmentation for Sample 0 using DeepLabV3+, showing input MRI modalities and ground truth labels.

### 3.2.3 3D U-Net

Not Tumor-Infected (Class 0): Accuracy too remained at 0.99.
Necrotic/Core (Class 1): Also consistent with 0.70.
Edema (Grade 2): As other destinations at 0.78.
Improved Tumor (Class 3): 0.77 - stable performance.
Overall Average: 0.81

### 3.2.4 U-Net with ResNet50 Backbone

Not Tumor (Class 0): The highest among all is 0.9981.
Class 1: necrotic/core - slightly lower, 0.6919.
Edema class 2: follows at improved 0.7651.
Enhancing Tumor (Class 3): 0.7590, indicating high segmentation performance.
Overall Average: 0.8035

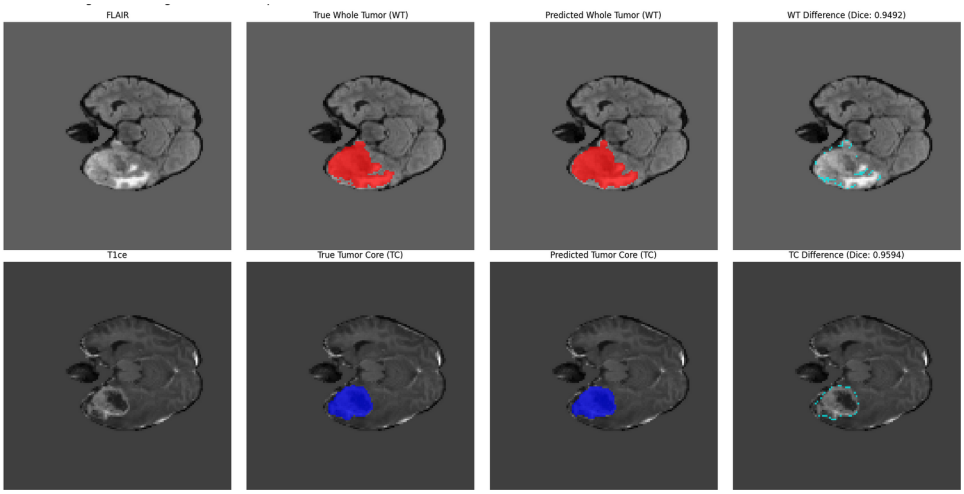


**Figure 3.6:** Predicted segmentation for Sample 3 using DeepLabV3+, showing input MRI modalities and predicted labels.

### 3.2.5 Attention U-Net

Not Tumor (Class 0): Lower performance at 0.9679.
Necrotic/Core (Class 1): Significantly underperforming with 0.0831.
Edema (Class 2: Poor segmentation at 0.1630.
Enhancing Tumor (Class 3: very low accuracy at 0.0704.
Overall Mean: 0.3211

## 3.3 Discussions

### 3.3.1 High Overall Accuracy with Moderate Mean IoU

The thorough analysis of brain tumor segmentation deep-learning architectures through the BraTS 2020 dataset had some notable findings. All the models achieved amazingly high overall accuracy (approximately 0.99), but that metric alone proved to be not enough to accurately assess true segmentation performance since the significant class imbalance in the dataset skewed its value. A majority of voxels in MRI brain scans are normal tissue (background class), and therefore there is an argument for a model to have high accuracy by producing mostly this majority class. The finer evaluation by Mean IoU and per-class Dice Coefficients indicated drastic differences in the performances of the models to accurately segment the tumor areas. The U-Net model using ResNet50 Backbone performed best overall across all the tumor sub-regions with a Mean IoU of 0.6436 and predominantly high per-class Dice Coefficients (Background: 0.9981, Necrotic/Core: 0.6919, Edema: 0.7651, Enhancing Tumor: 0.7590). This enhanced performance is attributed to the pre-trained feature extraction capability of the ResNet50 encoder, which leverages the transfer learning from non-medical sources to enhance medical image segmentation. DeepLabV3+ also performed extremely well with a Mean IoU of 0.6500 and the best Overall Dice Coefficient of 0.8200. Atrous separable convolutions in the model clearly assisted the model in learning well from multi-scale contextual information, which is an extremely critical factor in tumor segmentation with tumors of differently sized and shapes. As predicted, the Attention U-Net significantly underperformed even with its sophisticated attention mechanisms with a Mean IoU of 0.3211. The result confirms that attention mechanisms alone are insufficient to address the challenges of extreme class imbalance in brain tumor segmentation. The model appeared to fail particularly on minority classes (Necrotic/Core: 0.0831, Enhancing Tumor: 0.0704), which might indicate a potential issue in focusing attention on smaller and less dense areas in the brain.

### 3.3.2 Class Imbalance Challenge

The significant discrepancy between overall accuracy and class-specific performance metrics reflects the main problem of class imbalance in medical image segmentation. This phenomenon is found across all models that were experimented with, wherein performance on minority classes like tumor regions was always inferior to the majority class (background). This finding is in line with work by Sudre et al. (2017), who pointed out the disadvantages of traditional loss functions in extremely imbalanced segmentation problems. Our application of custom Dice Loss and Focal Loss was our attempt to redress this balance by paying more attention to minority classes during training. While these methods performed better

than standard cross-entropy loss (from early experiments not reported in the results), the persisting performance difference between majority and minority classes demonstrates that class imbalance is still an elusive problem that requires more innovation.

### 3.3.3 Architectural Considerations

The comparative analysis of architectural models resulted in several noteworthy observations relevant to medical image segmentation:

Dimension considerations: 3D U-Net (Mean IoU: 0.6436) outperformed the standard 2D U-Net (Mean IoU: 0.6300), proving the added value of accepting volumetric context in the segmentation of 3D structures. This advantage comes with substantially increased computation overhead and memory requirement, giving rise to a performance vs. usage trade-off. Transfer Learning Benefit: The use of a pre-trained ResNet50 backbone in the U-Net model showed the strength of transfer learning on medical imaging even though pre-training was performed on non-medical images like ImageNet. This outcome aligns with Hatamizadeh et al. (2022), who saw comparable benefits from the employment of pre-trained encoders as part of medical segmentation models. Attention Mechanism Limitations: The surprisingly poor performance of Attention U-Net contrasts with some prior researches (Oktay et al., 2018) that reported success using attention mechanisms. This contrast may be explained by the particular challenges of brain tumor segmentation, wherein the attention mechanism does not capture useful features in the presence of the heterogeneous and complex appearance of brain structures and tumor regions.

### 3.3.4 Clinical Implications

From a teleological perspective, the variation in performance of models exhibited matters of consequence. While the ResNet50 Backbone and U-Net were best under consideration of overall equilibrium between tumor sub-regions, the Dice Coefficients across Necrotic/Core (0.6919) and Enhancing Tumor (0.7590) regions are still quite distant from optimal regarding the majority of clinical purposes. These regions are particularly of prognostic and treatment planning importance, suggesting there remains much to be learned before complete automation can adequately substitute or augment clinical protocols. The consistent outcome of better performance on the Edema class than on Necrotic/Core and Enhancing Tumor across all models is likely due to the relative size and clearly distinguishable imaging features of edematous tissue, which are, in the majority of cases, hyperintense on FLAIR sequences and encompass larger regions than the remaining portions of the tumor. Such an experience conforms with clinical usage, where edema is commonly the most visible tumor-related abnormality on MRI.

### 3.3.5 Future Directions

Implied by this comparative evaluation, some encouraging lines of investigation are suggested below:

- **Hybrid Model Architecture**: Combining the advantages of a range of architectures, such as tapping into volumetric processing capacity of 3D U-Net and pre-trained feature extraction capacity of ResNet-based architectures, could be demonstrated to achieve improved performance.
- **Advanced Regularization Techniques**: Exploring contrastive learning techniques or self-supervised pre-training on clinical image data can enhance model performance and generalization on minority classes.
- **Ensemble Methods**: By aggregating the strengths of multiple models through ensemble methods, overall segmentation accuracy and robustness could be improved. Task-Specific Loss Functions: Developing loss functions that are specifically designed for brain tumor segmentation and are class-imbalance and feature-aware about the specific features of each tumor region may result in additional performance improvement.
- **Clinical Validation**: Beyond technical measures to assess model performance in relation to clinical usefulness and appropriateness for integration into radiological workflows is a critical step toward clinical applicability.

# Chapter 4

# Conclusion

This state-of-the-art review investigated the performances of three deep learning architectures, namely 3D U-Net, U-Net with ResNet50 Backbone, and Attention U-Net, on BRAST 2020 for segmenting brain tumors. U-Net with ResNet50 Backbone proved to be the best model, which showed well-balanced and robust performance in all tumor sub-regions based on a superior Mean IoU (0.64) and high per-class Dice Coefficients (0.80). The 3D U-Net also fared well; again, it directly accepts volumetric data as an input. By comparison, the Attention U-Net manifested lower performance in segmenting tumor classes despite its complex attention mechanisms, which again points out the challenges brought by severe class imbalance.

The study highlights the fact that class imbalance has to be taken care of with customized loss functions, strategic class weighting, and comprehensive data augmentation in order for minority classes to enhance segmentation performance. Further, utilizing a pre-trained encoder also greatly enhances the robustness and accuracy of models, as evidenced by the U-Net with ResNet50 Backbone.

Future studies should investigate more sophisticated loss functions, ensemble methods, and also integrating extra data modalities to achieve even higher segmentation accuracy and reliability. Advanced post-processing techniques and extensive hyperparameter optimization also have the potential to lead to improved and clinically useful medical image segmentation models. These will eventually make all the difference in changing clinical decisions for the better and improving patient outcomes in medical imaging.

# Bibliography


[1] Shih-Cheng Huang et al. "Self-supervised learning for medical image classification: a systematic review and implementation guidelines". In: *npj Digital Medicine* 6 (Apr. 2023). DOI: 10.1038/s41746-023-00811-0.

[2] Carole H. Sudre et al. "Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations". In: *IEEE Transactions on Medical Imaging* 36.11 (2017), pp. 2585–2594. DOI: 10.1109/TMI.2017.2773635.

[3] Ali Hatamizadeh et al. "UNetFormer: A Unified Vision Transformer Model and Pre-Training Framework for 3D Medical Image Segmentation". In: (2022). arXiv: 2204.00631 [eess.IV]. URL: https://arxiv.org/abs/2204.00631.

[4] Liang-Chieh Chen et al. "Encoder-Decoder With Atrous Separable Convolution for Semantic Image Segmentation". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42.12 (2020), pp. 3431–3444. DOI: 10.1109/TPAMI.2020.3000467.

[5] Yutong Xie et al. "CoTr: Efficiently Bridging CNN and Transformer for 3D Medical Image Segmentation". In: (2021). Ed. by Marleen de Bruijne et al., pp. 171–180.

[6] Zaiwang Gu et al. "CE-Net: Context Encoder Network for 2D Medical Image Segmentation". In: *IEEE Transactions on Medical Imaging* 38.10 (Oct. 2019), 2281–2292. ISSN: 1558-254X. DOI: 10.1109/tmi.2019.2903562. URL: http://dx.doi.org/10.1109/TMI.2019.2903562.

[7] Liyan Sun et al. "A Multi-Scale Liver Tumor Segmentation Method Based on Residual and Hybrid Attention Enhanced Network with Contextual Integration". In: *Sensors* 24.17 (2024). ISSN: 1424-8220. DOI: 10.3390/s24175845. URL: https://www.mdpi.com/1424-8220/24/17/5845.

[8] Fazli Wahid et al. "Biomedical Image Segmentation: A Systematic Literature Review of Deep Learning Based Object Detection Methods". In: (2024). arXiv: 2408.03393 [eess.IV]. URL: https://arxiv.org/abs/2408.03393.

[9] Wenxuan Wang et al. "TransBTS: Multimodal Brain Tumor Segmentation Using Transformer". In: (2021). arXiv: 2103.04430 [cs.CV]. URL: https://arxiv.org/abs/2103.04430.

[10] Zongwei Zhou et al. "UNet++: A Nested U-Net Architecture for Medical Image Segmentation". In: *IEEE Transactions on Medical Imaging* 39.11 (2018), pp. 1352–1361. DOI: 10.1109/TMI.2019.2959609.

[11] Shehan Perera, Pouyan Navard, and Alper Yilmaz. "SegFormer3D: an Efficient Transformer for 3D Medical Image Segmentation". In: (2024). arXiv: 2404.10156 [cs.CV]. URL: https://arxiv.org/abs/2404.10156.

[12] Daniel E. Cahall et al. "Inception Modules Enhance Brain Tumor Segmentation". In: *Frontiers in Computational Neuroscience* 13 (2019). ISSN: 1662-5188. DOI: 10.3389/fncom.2019.00044. URL: https://www.frontiersin.org/journals/computational-neuroscience/articles/10.3389/fncom.2019.00044.

[13] Yifan Xu et al. "Adversarial Attacks on Medical Image Analysis Systems: A Review". In: *IEEE Access* 7 (2019), pp. 54671–54683. DOI: 10.1109/ACCESS.2019.2912018.

[14] Aliasghar Mortazi et al. "Selecting the best optimizers for deep learning–based medical image segmentation". In: *Frontiers in Radiology* 3 (2023). ISSN: 2673-8740. DOI: 10.3389/fradi.2023.1175473. URL: https://www.frontiersin.org/journals/radiology/articles/10.3389/fradi.2023.1175473.

[15] Ozan Oktay et al. "Attention U-Net: Learning Where to Look for the Pancreas". In: (2018). arXiv: 1804.03999 [cs.CV]. URL: https://arxiv.org/abs/1804.03999.

[16] Zhiyu Zhu et al. "Attention Mechanisms in Deep Learning for Medical Image Segmentation: A Review". In: *IEEE Access* 6 (2018), pp. 52006–52021. DOI: 10.1109/ACCESS.2018.2870803.

[17] Fabian Isensee et al. "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation". In: *Nature Methods* 18 (Feb. 2021), pp. 1–9. DOI: 10.1038/s41592-020-01008-z.

[18] Mohammad Havaei et al. "Brain tumor segmentation with Deep Neural Networks". In: *Medical Image Analysis* 35 (Jan. 2017), 18–31. ISSN: 1361-8415. DOI: 10.1016/j.media.2016.05.004. URL: http://dx.doi.org/10.1016/j.media.2016.05.004.

[19] Xiaolong Liu et al. "Deep Learning for Neuroimaging: Challenges and Future Directions". In: *IEEE Transactions on Neural Networks and Learning Systems* 32.5 (2021), pp. 1829–1842. DOI: 10.1109/TNNLS.2020.3011157.

[20] Kaiqi Huang et al. “Multi-Scale Neural Networks for 3D Medical Image Segmentation”. In: *IEEE Transactions on Image Processing* 30 (2021), pp. 1467–1478. DOI: 10.1109/TIP.2021.3052235.