MULTI TASK DEEP LEARNING FRAMEWORK FOR SEGMENTATION AND MULTI CLASS CLASSIFICATION OF BRAIN TUMORS FROM MRI IMAGES

by

Zakia Zaman (200021302)
Tanzim Noor Tanmoy (200021338)
Afia Tahsin Chowdhury (200021358)

A Thesis Submitted to the Academic Faculty in Partial Fulfillment of the Requirements for the Degree of

BACHELOR OF SCIENCE IN ELECTRICAL AND ELECTRONICS ENGINEERING

Department of Electrical and Electronic Engineering
Islamic University of Technology (IUT)
Gazipur, Bangladesh

October 2025

MULTI TASK DEEP LEARNING FRAMEWORK FOR SEGMENTATION AND MULTI CLASS CLASSIFICATION OF BRAIN TUMORS FROM MRI IMAGES

Approved by:

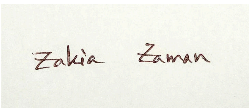
Ashraful Islam Mridha
Supervisor and Lecturer,
Department of Electrical and Electronic Engineering,
Islamic University of Technology (IUT),
Boardbazar, Gazipur-1704.

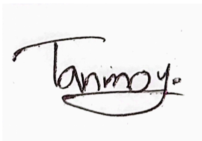
Date: 10/10/2025

Declaration of Authorship

This is to certify that the work presented in this thesis paper is the outcome of research carried out by the candidates under the supervision of **Ashraful Islam Mridha,** Lecturer, Department of Electrical and Electronic Engineering (EEE), Islamic University of Technology (IUT).

Authors





Zakia Zaman	Tanzim Noor Tanmoy	Afia Tahsin Chowdhury
Id : 200021302	Id : 200021338	Id : 200021358

Acknowledgment

First and foremost, Alhamdulillah. We feel incredibly grateful and indebted to the Almighty for guiding us through the journey of our thesis to completion. From the start of 4-1 to the end of 4-2, we faced numerous challenges- not only in our research but also beyond it. IUT was shut down twice, once due to a country-wide regime change and again due to a campus-wide reformation protest. Amid times of absolute uncertainty, we are truly grateful to have completed a project into which we poured so much effort and dedication.

We would like to express our deepest gratitude to our supervisor, **Ashraful Islam Mridha Sir**. Without his guidance, patience, and unwavering support, navigating this journey would have been impossible.

We are equally indebted to **Nadim Ahmed Sir**, whose combination of profound knowledge and humble demeanor is truly rare. His guidance was instrumental in shaping our approach, providing direction, and helping us transform our ideas into a tangible reality- so much so that we were able to publish our work in a conference long before our defense. Words cannot fully capture the significance of his mentorship and support.

A thesis, being a long-term endeavor, also requires the support of those beyond its direct scope. We are immensely grateful to our parents, siblings, and all others who supported us throughout this period, offering encouragement and strength in countless ways.

Finally, we want to acknowledge the friends and peers who surrounded us on campus. From answering random questions about the subject matter to helping with formatting or university formalities, their assistance and camaraderie were invaluable in bringing this work to life.

Abstract

Brain tumors are among one of the most critical and life threatening diseases that can lead to severe consequences or even death if it is not detected and treated at an early stage. Accurate , timely and faster diagnosis of brain tumor is crucial for improving the surviving rates of patients. However, throughout the years manual methods of brain tumor analysis has been prevalent where radiologists goes through a painstaking process of scanning through numerous MRI (Magnetic Resonance Imaging) images to identify the correct type of tumor/ This manual process is time consuming and often prone to human error. In order to address these challenges, deep learning based specially convolutional neural network based methods have gained increasing attention. Most conventional methods developed using deep learning architecture and convolutional neural network architecture mostly focuses one of the two aspects : either classification or semantic segmentation. In our work, we propose a multi-task deep learning framework that integrates both segmentation and multi class classification for comprehensive brain tumor analysis. The proposed framework employs Attention Residual U-Net model for precise tumor segmentation using the Figshare Brain Tumor Dataset, achieving a dice coefficient of 0.80 indicating strong overlap between predicted mask and ground truth tumor masks. The segmented tumor masks are subsequently concatenated with the corresponding MRI images to generate a two channel image highlighting the tumor region and providing classifier with spatial and contextual feature information. These images are then used to train an InceptionV3 classifier based multi class classification model to differentiate among three tumor categories meningioma, glioma and pituitary. The classification framework incorporated with the two channel image achieved an accuracy of 93.23%, outperforming the conventional classification models which were trained using only the raw MRI images. The integration of segmentation guided enhancement within the classification pipeline significantly improved the model's ability to focus on tumor regions, leading to better performance. Our proposed model serves as an end to end solution for automated brain tumor analysis, capable of performing both semantic segmentation to recognize tumor region as well as classify and identify the correct tumor type. This process provides a reliable and scalable approach suitable for real world implementation and clinical applications.

Table of Contents

List of Figures

List of Tables

List of Acronyms

MRI	Magnetic Resonance Imaging
CNN	Convolutional Neural Network
AGRes Unet	Attention Gate Residual U Net
ResNet	Residual Neural Network
ReLU	Rectified Linear Unit
ADAM	Adaptive Moment Estimation
DSC	Dice Similarity Coefficient
GELU	Gaussian Error Linear Unit
ROI	Region of Interest
IoU	Intersection over Union
MTL	Multi Task Learning

Chapter 1

1. Introduction

1.1 Background and Motivation

Brain tumors represent one of the most critical neurological disorders, where accurate diagnosis and delineation directly influence treatment planning and patient survival. Magnetic Resonance Imaging (MRI) serves as the preferred modality for visualizing intracranial abnormalities owing to its superior soft-tissue contrast. Traditional manual annotation by radiologists is, however, time-consuming and prone to subjective variability, motivating the use of automated deep-learning techniques. Over the past decade, convolutional neural networks have achieved state-of-the-art performance in brain-tumor segmentation and classification tasks [6]–[14]. Among them, the U-Net model proposed by Ronneberger et al. [6] has become the backbone of most segmentation pipelines, while CNN architectures such as ResNet and Inception V3 have excelled in tumor classification. Nevertheless, these approaches often function independently, leading to duplicated processing and a lack of cross-task knowledge sharing. Recent research trends emphasize multi-task learning (MTL) as a promising strategy to jointly optimize correlated objectives, improve generalization, and enhance interpretability [16], [17]. Integrating attention mechanisms and task coupling within MTL frameworks has shown particular potential in medical-image analysis, motivating the present study.

1.2 Problem Statement

Despite advances in deep learning and medical imaging, automated brain tumor segmentation and classification remain challenging due to variations in tumor shapes, sizes, and intensities. Most existing approaches focus on either segmentation or classification separately, limiting their effectiveness in real-world clinical scenarios. Furthermore, medical images, particularly MRI scans, are highly sensitive to distortions caused by deep learning architectures with excessive layers, as observed in our experiments with ResNet.

This project focuses on performing segmentation and classification altogether serially in a combined framework, utilizing:

1. **Semantic segmentation** using Attention Residual U-Net to accurately delineate tumor regions.

2. **Multi-class classification** to categorize tumors into meningioma, glioma, and pituitary types using Inception V3.

By leveraging the **Figshare dataset** containing **3,064 T1-weighted** contrast-enhanced MRI images with segmentation masks, this project aims to build a dual combined-model architecture that enhances diagnostic efficiency and accuracy. Existing literature lacks comprehensive solutions that combine segmentation and classification effectively on the same dataset, creating a research gap that this project seeks to address.

1.3 Objectives

The primary objectives of this project are as follows:

1. **Perform tumor segmentation**: Implement and optimize a U-Net-based model to segment brain tumors with high accuracy, using Dice score as a performance metric.
2. **Enhance segmentation through backbone variations**: Evaluate different U-Net encoders, including VGG16 and ResNet50, to determine their impact on segmentation accuracy.
3. **Develop a classification model**: Use a Inception V3 classification model to categorize brain tumors into meningioma, glioma, and pituitary tumor classes.
4. **Compare performance across architectures**: Experiment with various CNN architectures (InceptionV3, ConvNeXt Tiny, EfficientNetV2, ResNet, VGG16, DenseNet121, Xception, MobileNetV2) and compare their classification accuracy.
5. **Evaluate binary segmentation performance**: Implement tumor vs. non-tumor classification using segmentation masks.
6. **Optimize and validate the models**: Use performance metrics such as Dice score for segmentation and accuracy for classification to evaluate model effectiveness.

1.4 Scope and Limitations

Scope:

- This project focuses on deep learning-based semantic segmentation and classification of brain tumors using MRI images.
- The dataset used is the Figshare brain tumor dataset, containing 3,064 T1-weighted contrast-enhanced MRI images.

- The segmentation model is based on a modified U-Net and is trained from scratch on the Figshare dataset, while classification is performed using the Inception V3 model with transfer learning from ImageNet.
- The study evaluates different backbone architectures for segmentation and different CNN models for classification.

Limitations:

- The study is limited to T1-weighted contrast-enhanced MRI images, which may not generalize well to other imaging modalities.
- Segmentation and classification accuracy are subject to dataset quality and availability; more extensive datasets could further improve model performance.
- Computational constraints may limit hyperparameter tuning and training on deeper architectures.
- This project does not include clinical validation; real-world applicability needs further testing with clinical radiologists.

1.5 Literature Review

Recent advancements in deep learning have transformed automated brain-tumor analysis, particularly in magnetic resonance imaging (MRI)–based segmentation and classification. This section synthesizes prior single-task and multi-task learning (MTL) approaches, identifying key architectural innovations, datasets used, and remaining research gaps that motivated this study.

1.5.1 Single-Task Methods: Segmentation and Classification

Deep learning has significantly transformed MRI-based brain-tumor analysis, where early research typically approached segmentation and classification as independent single-task problems. Several studies [1]–[5] highlight that the U-Net architecture introduced by Ronneberger et al. [6] remains a foundational model for biomedical image segmentation because of its encoder–decoder structure with skip connections preserving spatial detail. Later variants such as Res-U-Net and Attention U-Net introduced residual mapping and attention mechanisms to improve gradient flow and concentrate on tumor-salient regions [7]–[9]. Benchmarks like the BraTS 2018–2020 and Figshare Brain-Tumor MRI datasets [10], [11] have been commonly used to evaluate segmentation models. Nevertheless, these architectures often lose fine spatial information in small lesions during down-sampling, reducing Dice performance [12].

For classification, CNN-based models such as VGGNet, ResNet, and Inception V3 achieved strong performance in distinguishing glioma, meningioma, and pituitary tumors [13], [14]. Despite their success, these networks generally depend on global image features without explicit spatial tumor context, which limits interpretability from a clinical perspective [15].

1.5.2 Emergence of Multi-Task Learning in Medical Imaging

To overcome the limitations of isolated single-task pipelines, researchers have employed multi-task learning (MTL), which jointly optimizes segmentation and classification tasks through shared representations. MTL enables more efficient learning from limited annotated data, enhances feature reuse, and improves generalization performance [16], [17]. In brain-tumor MRI analysis, MTL models are broadly divided into **parallel** and **serial** frameworks depending on how the tasks interact within the network.

1.5.3 Parallel Multi-Task Learning Architectures

Parallel MTL architectures share a common encoder that extracts mutual features, followed by task-specific decoders or heads for segmentation and classification. Nazir et al. [19] proposed a 2D MTL framework that simultaneously performs tumor detection and segmentation using a shared convolutional backbone, demonstrating improved convergence stability. Gupta et al. [20] developed MAG-Net, a multi-task attention-guided network employing a unified encoder to drive both segmentation and classification branches, achieving robust results on the Figshare dataset. Similarly, Huang et al. [21] introduced a V-Net-based MTL model that jointly optimizes segmentation and auxiliary distance-map prediction, which enhances contour precision and boundary regularity. These parallel approaches emphasize shared feature learning and efficient utilization of global context; however, they may still struggle when the classification task requires localized tumor information that differs from the segmentation objective.

1.5.4 Serial Multi-Task Learning Architectures

In contrast, serial MTL frameworks design a hierarchical task dependency in which the output of one task directly informs the next. Sobhaninia et al. [22] proposed a cascaded model that first performs tumor localization, then refines segmentation and classification in subsequent stages. Their design ensures that the classifier operates only on the relevant tumor area rather than the entire MRI slice. Rabby et al. [23] advanced this concept by integrating the segmentation mask as an auxiliary input for classification, allowing the classifier to focus spatially on diagnostically significant regions. Such serial coupling enhances interpretability and spatial awareness and often yields superior classification accuracy compared to parallel models because it explicitly connects tumor structure with diagnostic outcome.

1.5.5 Attention Mechanisms and Feature Modulation

Attention mechanisms have become central to modern medical-image architectures. Oktay et al. [24] proposed the Attention U-Net, where gating modules embedded in skip connections adaptively emphasize salient spatial features and suppress irrelevant background. Building upon this idea, Zhang et al. [25] developed the Attention-Gate Residual U-Net (AGResU-Net), combining residual learning with attention gates to strengthen gradient propagation and improve segmentation accuracy, especially for small or diffuse tumors in the BraTS and Figshare datasets [10]. Ngo et al. [26] later demonstrated that augmenting attention-based MTL models with feature-reconstruction objectives enhances the network's ability to detect small or low-contrast tumor areas by preserving fine spatial information.

1.5.6 Hybrid and Enhanced Segmentation Models

Beyond classical U-Net derivatives, several hybrid architectures have sought to capture broader spatial dependencies. Li et al. [27] proposed an enhanced 3D U-Net integrating modified encoder–decoder modules and a hybrid cross-entropy–Dice loss, yielding higher Dice coefficients for core and enhancing tumor regions on BraTS data. Qamar et al. [28] introduced the Hyperdense Inception 3D U-Net (HI-Net), which combines inception-style and residual blocks to extract multi-scale spatial features efficiently. Nazir et al. [19] and Gupta et al. [20] extended similar design principles to multi-task frameworks, while Lv et al. [29] proposed BrainTumNet—a unified multi-task deep network performing both segmentation and classification concurrently. Odong et al. [30] also presented an integrated CNN capable of executing both operations simultaneously, improving interpretability while reducing redundant computation.

1.5.7 Research Gaps and Positioning of the Present Study

Despite remarkable progress, several limitations remain evident across the literature. Many frameworks still operate as disjointed or loosely coupled pipelines, preventing optimal synergy between segmentation and classification tasks [19]–[21]. Standard U-Net variants often lose contextual information through down-sampling, impairing detection of small lesions [8], [25]. Heavy reliance on BraTS datasets restricts generalization, with limited testing on alternate datasets such as Figshare [10], [23]. Additionally, large encoder backbones like ResNet-50 increase inference latency, posing challenges for real-time clinical use [14]. Finally, many systems lack sufficient interpretability for practical deployment.

The present study directly addresses these shortcomings by designing a serial multi-task deep-learning framework coupling an Attention Residual U-Net-based segmentation backbone with an Inception V3 classifier. Here, the segmentation output acts as a spatial guide to the classifier, enabling the model to focus on tumor-specific regions and improving diagnostic reliability. Evaluations on the Figshare dataset achieved a Dice coefficient of 0.80 for segmentation and 93.23 % classification accuracy, validating the effectiveness of a unified, interpretable, and computationally efficient approach to brain-tumor MRI analysis.

Chapter 2

2. Methodology

The proposed research presents a multi-task deep learning framework designed to perform both semantic segmentation and multi-class classification of brain tumors from Magnetic Resonance Imaging (MRI) scans. The overall methodology is structured into two major components:

- Segmentation of tumor regions
- Classification of tumor type

This two-stage architecture was developed to provide an end-to-end automated pipeline capable of identifying the precise tumor region and subsequently classifying it into distinct categories: meningioma, glioma, or pituitary tumor. The motivation for designing a multi-task framework lies in the clinical importance of both identifying and categorizing brain tumors accurately, as segmentation alone indicates the tumor's location but not its type, while classification without segmentation may overlook precise spatial information essential for diagnosis.

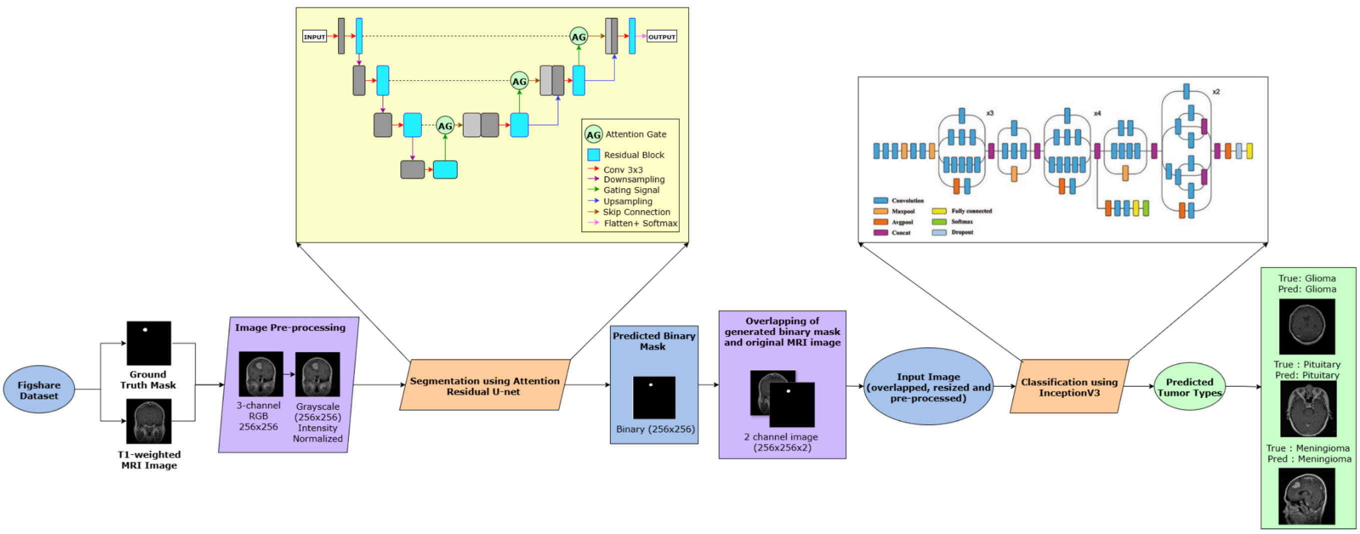


Fig 2.1.1: Workflow of Proposed Framework

2.1 Dataset

The chosen dataset for our work is the Figshare brain tumor dataset.Our dataset contains 3064 T1 weighted contrast enhanced MRI images from 233 patients with categorized in three types of brain tumor. The three kinds of brain tumor are meningioma (708 slices), glioma (1426 slices) and pituitary tumor (930 slices). The dataset is divided into four subsets and archived in zip format. The archived data contains label for three different types of tumor, the patient Id, Image data in ".mat" format, tumor border and binary tumor mask.

The data was extracted from .mat format using h5py library and saved as NumPy arrays. Extracted data was then normalized and resized for fitting in different deep learning framework based on which model was being used to train the data.

2.2 Semantic Segmentation

2.2.1 Baseline Model : U Net

The architecture of Unet is based on the encoder-decoder structure. One contracting path and once expanding path. The purpose of the contracting path is to decrease the spatial dimension of the image and at the same time capture the feature information by generating feature maps in different layers. The expanding path targets to upsample the feature map and produce segmentation map by utilizing the features learnt from the contracting path. The contracting path in the U-net consists of convolution as well as pooling layers. The model takes a 2D image as input of size 256 by 256. The kernel size used here is 3 by 3. Every convolutional layer is followed by a max pooling layer which downsamples the feature map to half its size. This procedure is repeated for five blocks along the contraction path. The expanding path uses both convolution and up-convolution layer to combine learnt features and upsample the input feature map until a segmentation map is generated. The final layer retains the initial size of the image and gives a binary segmentation map.

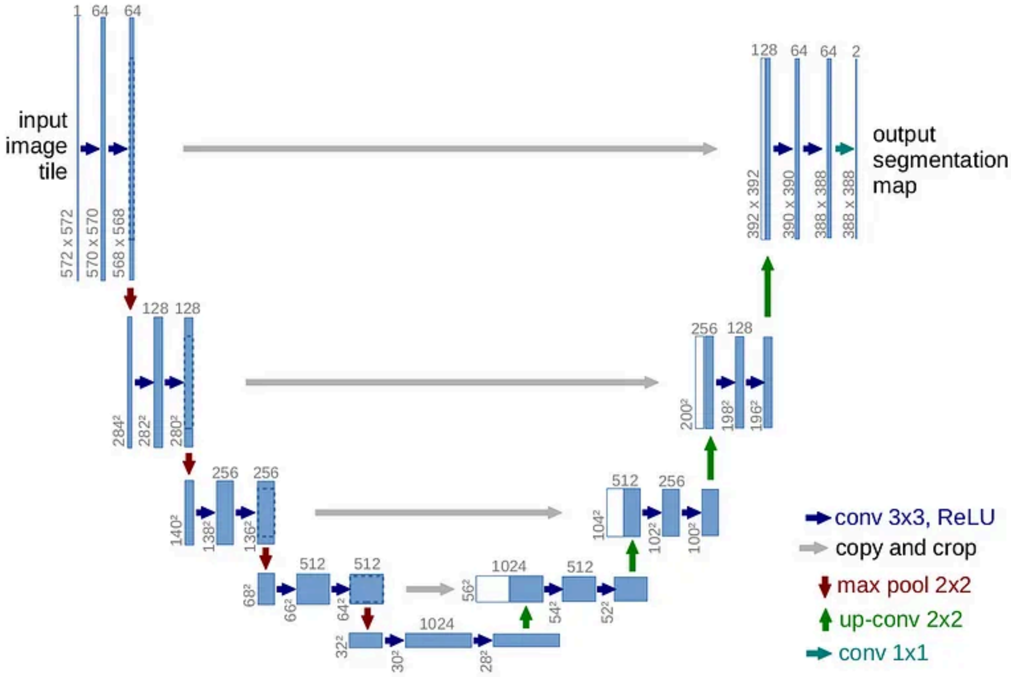


Fig 2.2.1: Baseline U Net Model Architecture (Ronneberger et al., 2015)

2.2.2 U Net with Shared Encoders

U Net with VGG16 as Encoder

The VGG16 backbone was adapted into the U Net architecture by changing the native encoder block with the pre trained VGG16 model using its hierarchical feature extraction capabilities for tumor segmentation. The VGG16 encoder, pre-trained ImageNet dataset consists of 13 convolutional layers grouped into five blocks, each followed by max pooling (2x2 stride) to reduce spatial dimensions and at the same time to increase feature depth. To integrate the model into U Net the fully connected layers were discarded and its convolutional blocks were repurposed as the contracting path. Skip connections were established between the outputs of VGG16's pooling layers (Block1- Block5) and the decoder's upsampling path to preserve spatial details. However, adjustments were made to align feature map dimensions: the decoder used transposed convolutions to upsample feature maps, concatenating them with corresponding skip connections from VGG16 before applying sequential convolutions (3×3 kernels, ReLU activation). Blocks 4–5 were fine-tuned to adapt to medical imaging patterns. Despite VGG16's robust feature hierarchy, its large parameter count (138 million) and fixed kernel initializations (optimized for natural RGB images) introduced challenges, such as over-regularization on the smaller medical dataset and domain mismatch between natural and grayscale MRI images. This resulted in suboptimal skip connection fusion, ultimately yielding a lower Dice score (70%) compared to the original U-Net (71%), highlighting the trade-off between transfer learning benefits and architectural compatibility in medical image segmentation tasks.

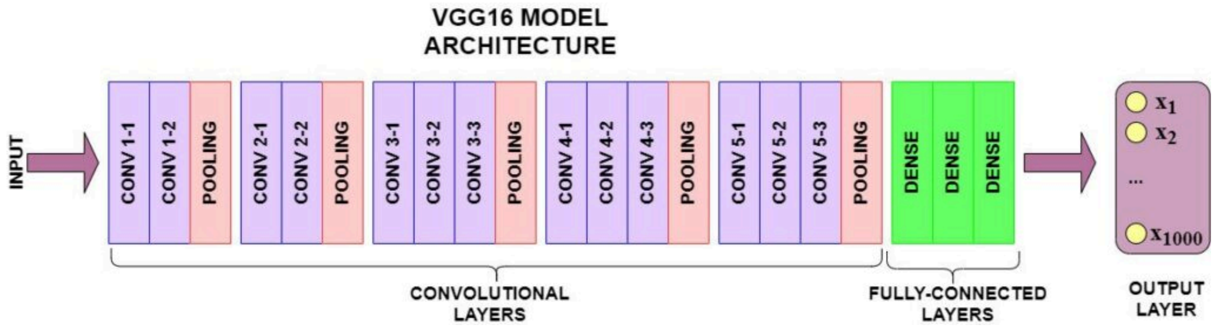


Fig 2.2.2: VGG16 Model Architecture (McDermott, n.d.)

U Net with ResNet50 as Encoder

The ResNet50 backbone was incorporated into the U-Net architecture by replacing its native encoder with a pre-trained ResNet50 model, using its residual learning framework to enhance gradient flow during tumor segmentation. ResNet50's encoder consists of 49 convolutional layers organized into 4 residual blocks (conv2_x to conv5_x), each containing multiple bottleneck layers with identity skip connections to mitigate vanishing gradients. The original fully connected layers were discarded, and the encoder outputs were extracted at four hierarchical stages:

- `conv1_relu`
- `conv2_block3_out`
- `conv3_block4_out`
- `conv4_block6_out`

The decoder path uses transposed convolutions to upsample feature maps, concatenating them with corresponding skip connections from ResNet50's encoder blocks via decoder_block. To address channel depth mismatches between ResNet50's output (2048 channels at the bridge) and the decoder's expected dimensions, the first transposed convolution in each decoder_block reduces channels before concatenation. Despite ResNet50's superior performance on natural images (achieving 65% Dice score vs. VGG16's 70%), its residual blocks introduced spatial incompatibility in skip connections due to aggressive downsampling (stride=2 in residual blocks), causing misalignment between encoder and decoder feature maps. Furthermore, the model's 25.6M parameters increased computational costs compared to the native pretrained U-Net (7.8M).

2.2.3 Attention U Net

Fully convolutional architectures and U Net architectures perform fairly well in brain tumor image segmentation. However, models trained with attention gates learn to emphasis on the relevant regions and disregard the less important areas of the input image. This process of emphasizing, highlights the salient features useful for specific task like segmentation of tumor from MRI images. The problem with native U Net based architectures was that the features at lower stages are repeatedly extracted and due to the skip connections the high level feature maps tend to get distorted. This process also causes excessive use of computational resources. This is where the benefit of attention networks come in. The attention gates generate soft region based attention implicitly during training and emphasis on the salient features for the tumor segmentation.

The architecture follows the same encoder-bottleneck-decoder method of the native U Net. Here the encoder blocks are consisted of convolutional plus max pooling operations to capture features at different spatial levels. The decoder upsamples and performs convolution operation to reconstruct the spatial resolution. Now the difference with traditional U Net architecture lies in the skip connections. In U Net the skip connections directly pass encoder features to the decoder. In Attention U Net each skip connection passes through an attention gate, which filters the encoder feature and highlights the salient features before concatenation.

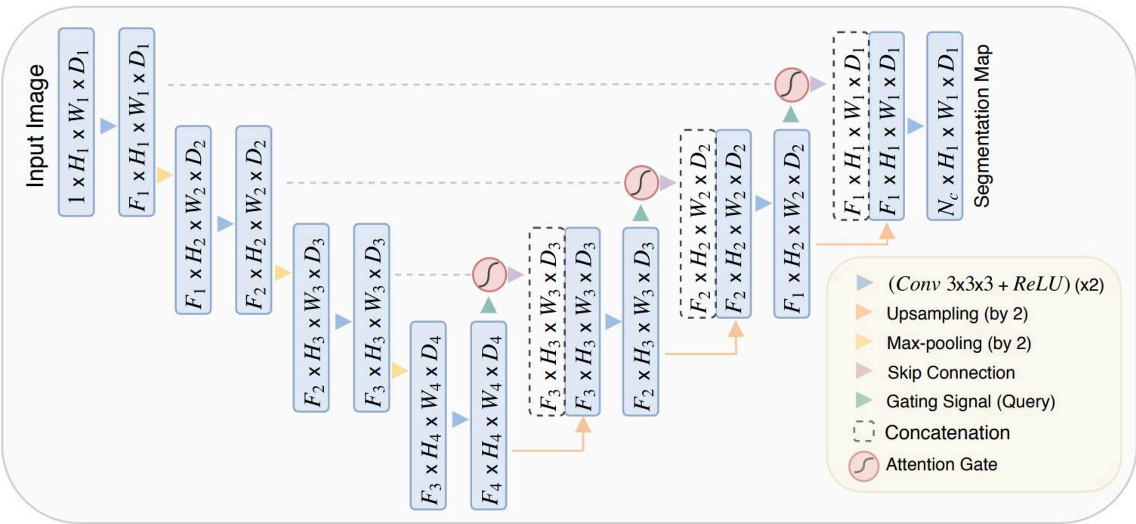


Fig 2.2.3: Attention U Net Architecture (Otkay et al.)

Attention Gate : An attention gate is a trainable module that learns to weight the importance of features spatially and it takes two inputs. The two inputs are gating signal (g) and features from the encoder side (x). The Attention Gate architecture sums the input and gating signal and pass them to a

ReLU activation function which generates intermediate feature map. This feature map is passed through a convolution layer and after that passed on to a sigmoid activation layer to produce a final attention coefficients. The attention coefficients specify the salient image regions and prune feature responses to preserve only the activations relevant to the brain tumor segmentation. Hence, it focuses only on the tumor region.

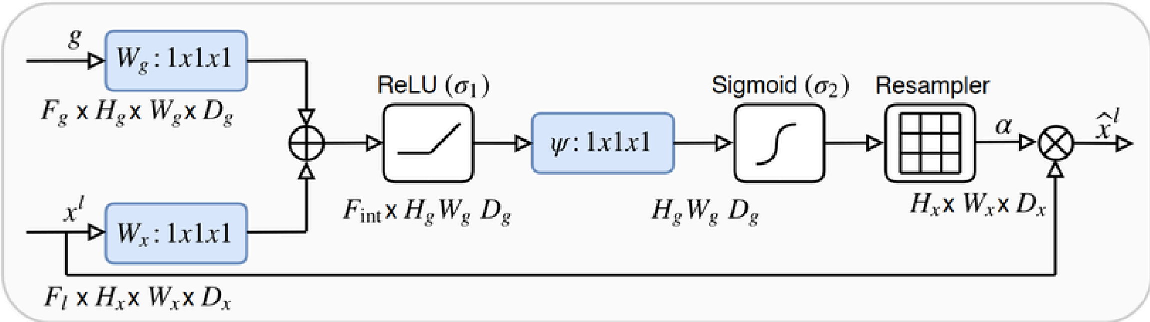


Fig 2.2.4: Attention Gate Architecture (Otkay et al.)

The selective filtering from the attention gates ensure that only relevant spatial regions are passed on to the decoder so that unnecessary information are filtered out. This model gave a better result than baseline U Net models and the test dice score was 76%.

2.2.4 Residual Attention U Net

The previous model of attention U Net has been improved upon by incorporating residual blocks in the Attention Gate Residual U Net model. This model integrates residual blocks as well as attention gates with the native U Net model in which the attention gates are used in the skip connections for highlighting the salient feature information while discarding unwanted and noisy features. In this improved version of the model the residual attention gate unet not only extracts the semantic information to increase the feature importance but also learns the small scale information of brain tumors. Due to the complex structure and diverse dimensional changes of brain tumors, very small scale features exist which might get lost in the deeper stage of downsampling in native unet based architectures. The spatial information and location details of the high level output feature maps gets jumbled up in the procedure of cascading and downsampling. The attention gate solves this issue by propagating only the relevant tumor region information and spatial details of low level feature maps in the skip connection process. For more promising segmentation performance the attention gates are integrated with the residual blocks to produce Attention Gate Residual U Net.

The architecture follows the original U Net model and consists of encoder, bottleneck and decoder. The difference lies that after every convolution block a residual block is added. This residual block adds a shortcut path to allow a gradient flow directly through the network for concatenating the previous features with the newer features. In this way the model learns residual mapping instead of direct transformations.

Next attention gates are added at the skip connections where the gates take two inputs. One is the gating signal and the other one is the skip connection connecting the encoder side. The gating signal is originated from the decoder side. Here we utilize 3x3 convolutional kernel with a stride of 2 for the downsampling. The contracting path consists of three residual blocks and each residual block has a convolutional block and a max pooling layer. The ReLU activation function is used for the convolutional block while the final activation function for the model is Sigmoid as we will be having a binary segmentation map as output.

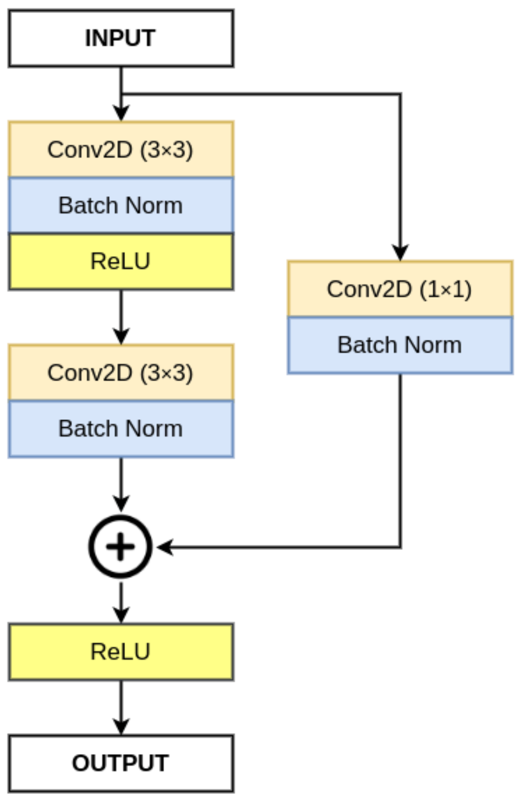


Fig 2.2.5 : Residual Block Architecture [31]

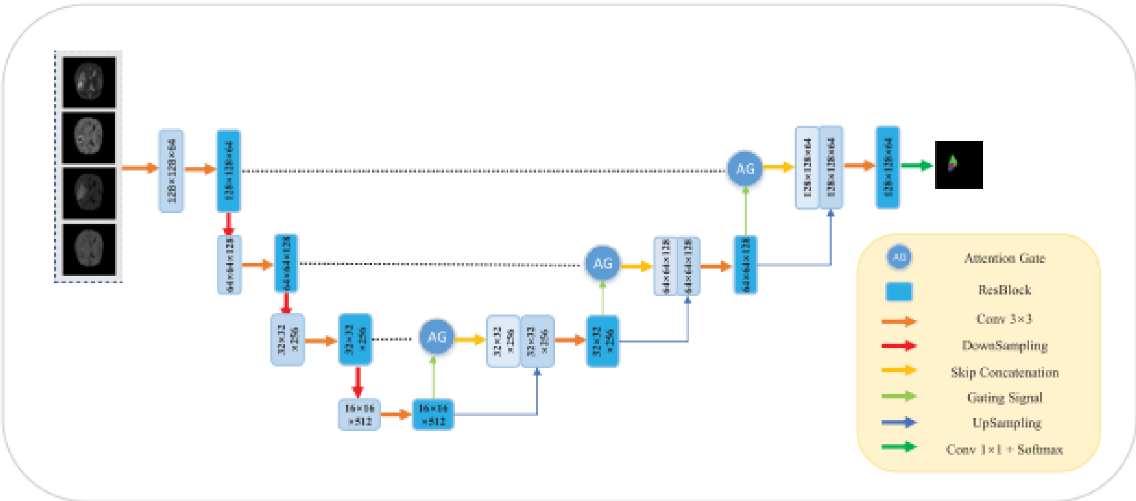


Fig 2.2.6: Attention Residual U Net Architecture (Zhang et al.)

2.2.5 Loss Function and Optimizer

The dice coefficient and dice score were chosen as the loss function for smeantic segmentation. Here the dice loss function was used to optimize the tumor segmentation performance, addressing the class imbalance inherent in medical imaging where tumor pixels are sparse compared to the background. The Dice Coefficient (DSC) measures mask overlap between the predicted (ypred) and ground truth (ytrue) masks and it is defined as ,

$$DSC = \frac{2.\sum(ytrue.ypred) + e}{\sum ytrue + \sum ypred + e}$$

Here e is the smoothing factor which prevents the division by zero. The custom dice_coef function flattens both masks into 1D vectors to compute pixel wise agreement while dice loss calculates the inverse of this metric denoted as ,

$$Dice\ Loss = 1 - Dice\ Coefficient$$

The optimizer for the model was chosen as the ADAM (Adaptive Moment Estimation) optimizer. This optimizer combines the best properties of AdaGrad and RMSProp. This optimizer is considered to be best because it adjusts learning rate for each parameter individually. Keeps track of the moving average of past gradient by calculating first moment and squared gradients from second moment. This optimizer outperforms other optimizers and converges the model really well.

2.2.6 Model Workflow

The segmentation workflow begins with dataset preparation. In this model we have used the benchmark Figshare Brain Tumor Dataset with 3064 T1 weighted contrast enhanced MRI images and corresponding binary tumor masks. The images and masks are extracted from the .mat format files and saved in a dataset containing two folders images and masks. The next step is data preprocessing which includes image resizing , normalization and augmentation. The images and masks are resized and cropped to 256 by 256 and then normalized between 0 and 1 by dividing the pixel values by 255. Train test split is used to divide the dataset in training data and testing data in 80:20 ratio. From the images and masks tensorflow datasets are generated for efficient training. In the next part, a custom attention residual U net architecture is designed which includes following key components. A convolution block which implements the convolution layer followed by batch normalization and ReLU activation layer. Residual block is added to the convolutional layer . The attention gate computes the gating and input feature maps and is applied on the skip connections.

A custom dice loss function and dice coefficient is defined. Dice loss and dice coefficient is the most common method for measuring the accuracy of the segmentation model. The model is compiled using Adam as the optimizer, dice loss as the loss function and the performance metric as accuracy and dice coefficient. The model was trained for 100 epochs with a learning rate 0.00001. Tensorflow callback functions were used for regularization and optimized training. The best model was saved during training using the model checkpoint and learning rate scheduler was also employed for better performance. The model performance was measured on the test dataset. Training accuracy and validation accuracy curve was observed for understanding the output of the model. Training loss and validation loss curve was also observed for better understanding of the model and observe whether the model has been trained properly or not.

2.3 Classification

2.3.1 Image Preprocessing for Multi Class Classification

After the semantic segmentation was performed the MRI images were further prepared for multi class classification model training. In regular classification model training only the image and its corresponding label is given as input to the model which the model is trained to recognize. The model learns patterns associated with the image such as the texture, border, size , area and is able to distinguish between different classes. In case of brain tumor classification using MRI images the model sees patterns of pixel intensities. Over many training examples it learns different statistical patterns corresponding to each tumor type. Each MRI image is generally given as a 256 by 256 grayscale or three channel RGB image. The model applies a convolutional filter on each pixel to detect local patterns and relation between pixels. In early layers, it derives low level features such as edges, contours and textures. In deeper layers and higher level of convolution the model learns more defining features like tumor location, size, shape and creates feature map in each layer.

As the aim of our research is to enhance the accuracy of classification alongside performing semantic segmentation we have incorporated the previously obtained predicted segmentation mask with the MRI images. Instead of giving only the MRI tumor image and classification labels as input to the classification model the MRI images were prepared as two channel input consisting of the original image and its predicted tumor mask. The best performing model from segmentation was chosen, which was the Attention Residual Unet model and the predictions from the model was saved as binary tumor masks. The corresponding image and masks are resized to 256 by 256 pixels and normalized. The normalized grayscale MRI image and mask were concatenated along the channel axis to form a two channel tensor. This two channel tensor was given as input alongside the classification labels to the model.

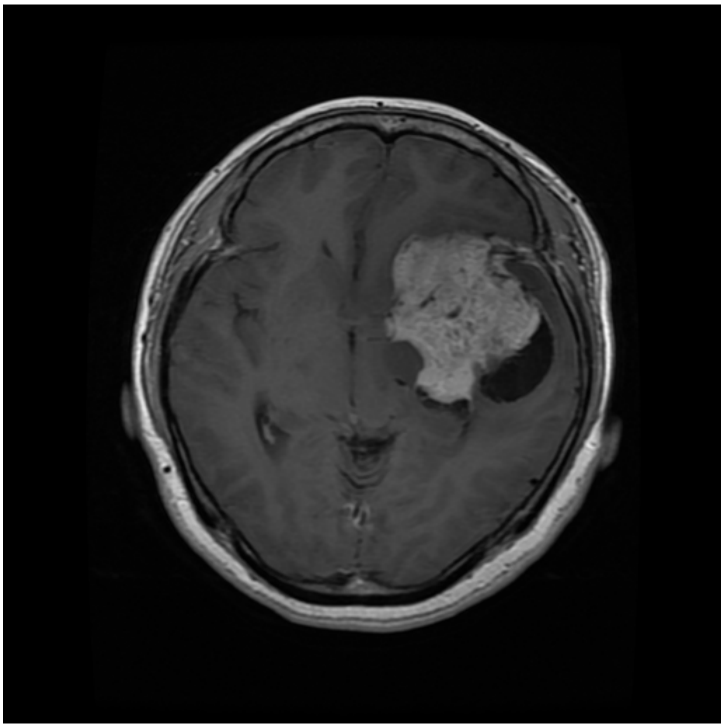


Fig 2.3.1: Example MRI image of brain

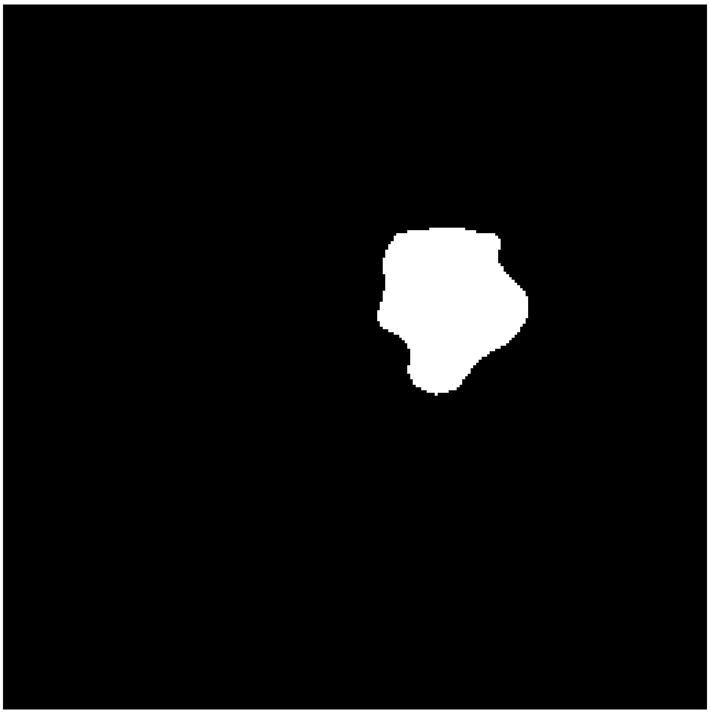


Fig 2.3.2: Corresponding Segmented Mask from Segmentation model output

By combining the image and mask, two complementary information is provided to the model which contain raw intensity, contextual information of tumor and anatomical details as well as precise tumor location, shape and region of interest (ROI). The mask works like a spatial attention map and removes unnecessary emphasis on background information. With this modification, the model learns features more efficiently and reduces overfitting. Finally after creating the dataset, it was split for training and testing.

2.3.2 InceptionV3 Architecture

The classification backbone adopted is InceptionV3, a deep convolutional neural network architecture. It is specifically designed to achieve high classification accuracy while maintaining computational efficiency. The InceptionV3 network consists of approximately 48 layers and about 25 million trainable parameters, offering a strong balance between depth and performance.

At its core, the InceptionV3 architecture is composed of a series of inception modules, each performing parallel convolutional operations with multiple kernel sizes such as 1×1, 3×3, and 5×5. The outputs of these parallel branches are concatenated along the channel axis, enabling the network to capture multi-scale spatial information from the input data. This design drastically reduces computational cost while preserving the model's ability to extract complex spatial features.

A distinctive characteristic of InceptionV3 is the use of factorized convolutions, where larger convolutions are decomposed into multiple smaller ones. This reduces the number of parameters, lowers the risk of overfitting, and accelerates the learning process without compromising representational power. The network also integrates batch normalization layers after each convolution to stabilize gradient flow and improve convergence.

Another important aspect of the InceptionV3 design is the inclusion of auxiliary classifiers in intermediate layers, which serve as regularizers by providing additional gradient feedback during training. The architecture concludes with global average pooling, followed by a fully connected dense layer and a softmax output layer, which produces the final class probabilities.

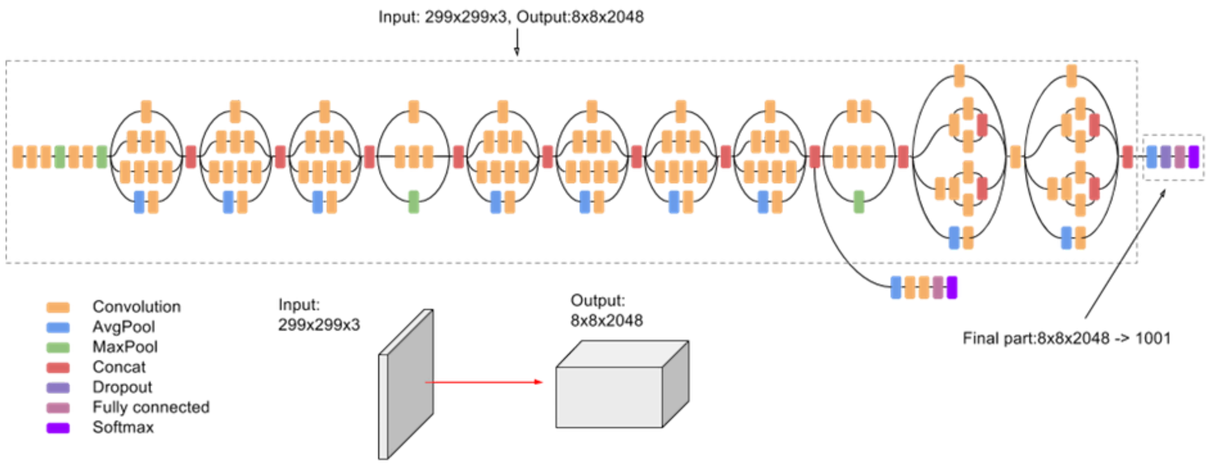


Fig 2.3.3: InceptionV3 Model Architecture [32]

In our implementation, the pre-trained InceptionV3 network was fine-tuned using the constructed brain MRI dataset. Since InceptionV3 expects a 3-channel input, the two-channel MRI-mask tensor was first expanded to three channels through an additional convolutional projection layer before entering the network.

The InceptionV3 backbone extracts discriminative features from the combined input by leveraging both texture-level cues from MRI intensities and spatial localization information from the predicted tumor mask. The final dense layer was modified to output three classes: meningioma, glioma, and pituitary tumors.

The selection of InceptionV3 as the classifier backbone was motivated by its multi-scale feature learning ability, computational efficiency, and robust generalization performance in medical image analysis. Compared to conventional CNNs, its inception modules allow it to simultaneously consider local and global patterns, making it well-suited for identifying subtle structural variations between different tumor types.

2.3.3 Classification Workflow

After preprocessing, we assemble the classification dataset by pairing each two-channel image tensor with its corresponding tumor label. We randomly split this dataset into training and test subsets so that model performance can be evaluated on unseen data as well. The labels are converted into one-hot vectors over the tumor classes, namely glioma, meningioma and pituitary. Each image-mask pair is normalized and resized to the network's input size to ensure that during training the model sees properly scaled inputs. In a CNN, successive layers learn hierarchical features: early layers detect simple edges or textures, while deeper layers combine these into more complex shapes.

The classifier backbone we used is InceptionV3. Before feeding the data into the classification network, the two-channel tensor is expanded to three channels through an additional convolutional operation. This step is necessary because the InceptionV3 backbone accepts a 3-channel input similar to RGB images. The expansion layer learns to project the two input channels into a compatible 3-channel feature representation without losing the spatial or contextual information provided by the mask. The classifier then processes input images through cascades of convolution and pooling layers that extract increasingly abstract features. The Inception architecture builds inception modules that apply several convolutional filters in parallel and concatenate their outputs, enabling the network to capture multi-scale spatial patterns efficiently. By combining raw image intensities and tumor-region information from the mask, the model simultaneously learns from both anatomical texture and explicit tumor localization.

Training is carried out for a fixed number of epochs with early stopping based on validation loss to prevent overfitting. Built-in dropout layers and batch normalization within the InceptionV3 architecture further regularize the model. The network's final layer is a fully-connected softmax that outputs class probabilities corresponding to the three tumor types.

Incorporating the segmentation mask greatly aids classification as the mask functions as a spatial attention map that highlights the tumor region and suppresses irrelevant background. By providing this auxiliary channel, the network focuses on tumor-specific features such as shape, boundary, and texture rather than normal anatomy.

2.3.4 Training of Standalone CNN Classifiers

To evaluate the effectiveness of the proposed multi-task framework and better understand its advantages, several convolutional neural network (CNN) architectures were also trained and tested on the Figshare Brain Tumor dataset. A total of 3,024 MRI images were utilized, consisting of 708 meningioma, 1,426 glioma and 930 pituitary slices.

Five widely recognized CNN architectures were selected for this comparative study: Inception V3, VGG16, VGG19, InceptionV3, EfficientNetV2S and ConvNextTiny. These models were chosen for their strong performance and popularity in the medical imaging domain. The MRI slices were preprocessed through resizing, normalization and conversion to RGB format to match the input specifications of each model.

Each classifier was trained using an 80:20 training and testing split and evaluated through five-fold cross-validation to ensure generalization and reduce bias. All models employed categorical cross-entropy loss with the Adam optimizer for optimization.

The results from these standalone classifiers were later compared with those of the proposed segmentation-guided InceptionV3 framework. This comparison clearly demonstrated that incorporating tumor masks as an auxiliary input substantially improved classification accuracy, precision, and robustness compared to the independently trained CNN models.

2.4 Model Optimization

Optimization refers to the process of minimizing the loss function and calculating the most optimal set of parameters for the model that gives the best result on a given task. During training, the model makes predictions, compares them to the true labels using a loss function and then adjusts its parameters through backpropagation using an optimization algorithm like Adam. Different optimization techniques are employed to work alongside the optimization algorithm to tackle different issues while training. The goal of optimization is to guide the model toward an optimal point in the loss function where it can make most accurate predictions without overfitting or underfitting the data or without any bias. In our work, we have used techniques like L2 regularization, label smoothing, class weight balance and cosine decay learning rate scheduler to tackle different challenges like overfitting, lack of generalization, class imbalance in training. Using different optimization technique provides smooth learning resulting in better generalization, stable training process and higher classification accuracy.

2.4.1 Class Weight Balance

In classification problem, the dataset is often unbalanced and contains unequal number of samples for each classes which causes disparity in training. The figshare dataset includes MRI images of three types of brain tumor which are meningioma, glioma and pituitary. As the number of sampless for each class is not balanced and there are less number of meningioma tumor images than the other two types the model might show bias towards the majority class which may lead to worse performance on the underrepresented one. In order to address this issue, the concept of customized class weight balance function was used.

Class weight balance is a method that adjusts how much importance the model gives to each class during training. In case of our dataset if one class has more samples than the other then the model naturally learns to favor the larger class as it sees it more often. By giving higher weights to the minority classes, the training process penalizes the mistakes made on rare classes and thus prevent misclassification of bias in model training. In the figshare brain tumor dataset there are 1426 glioma images, 930 pituitary images and 708 meningioma tumor images. Glioma images are most frequent while meningioma images are comparatively fewer. Without penalizing the model for biasing the model may predict glioma most of the time as it will learn the feature of the majority class more than the minority classes. Minority classes like meningioma can be ignored resulting in misclassification and low recall and precision score for those categories.

In our thesis we have utilized the “compute_class_weight” function from the “sklearn.utils.class_weight” library for the purpose of computing class weights for each classes. The following code snipped exhibits the implementation of the module.

```
from sklearn.utils.class_weight import compute_class_weight
import numpy as np


classes = np.unique(labels)
class_weights = compute_class_weight(class_weight='balanced', classes=classes, y=labels)


class_weights_dict = dict(zip(classes, class_weights))

print("Computed class weights:", class_weights_dict)
```

```
Computed class weights: {0: 1.4425612052730696, 1: 0.7162225338943432, 2: 1.0982078853046595}
```

Fig 2.4.1 : Code Snippet for Class Weight Balance Function

The function uses this following formula:

$$w_j = \frac{n_{samples}}{n_{classes} \times n_j}$$

where,

w_j = weight assigned to class j

$n_{samples}$= total number of samples

$n_{classes}$ = number of distinct classes

n_j= number of samples belonging to class j

According to this formula, classes with larger number of samples are assigned lower priority while the classes with smaller number of samples are assigned lower priorities. The computed class weights are saved as a dictionary that maps each class label to its corresponding weight. This dictionary is later passed on to the model while training in the training function "model.fit" ensuring that the loss function incorporates these weights during gradient updates. The main advantage of using this optimization method is that it directly addresses the issues of data imbalance and tackles the problem of misclassification. Thus avoiding bias towards glioma type and ensuring a balanced training across all the classes.

2.4.2 L2 Regularization

In deep learning models, particularly in medical image analysis, achieving high training accuracy does not always guarantee good performance in real world scenarios. Models trained on complex datasets like MRI brain tumor images are prone to overfitting where they learn the training data very precisely, resulting in a higher training accuracy but performs very poorly in unseen data. The models lack generalization. In order to mitigate this problem, L2 regularization has been used which is also known as weight decay. In our work, the L2 regularization method has been incorporated in the classification head of the model to enhance generalization capability of the model during training. The L2 regularization uses the following formula :

$$L_{total} = L_{original} + \lambda \Sigma w_i^2$$

where ,

$L_{original}$ = model's categorical cross-entropy loss

w_i = trainable weight parameter in the network

λ = regularization coefficient

Here, the regularization coefficient controls the strength of the penalty and it is set to 0.001 during training. By minimizing the total loss, the optimization process encourages the model to maintain smaller weight values. This results in smoother and simpler decision boundaries. As the dataset that we have used is rather small, contains less number of samples and is imbalanced, the risk of overfitting is high. L2 regularization constraints model's parameters and prevents it from relying too heavily on a few dominant features or training examples. The regularization is applied with a strength of 0.001 to the weights of the dense layer containing 256 neurons. During backpropagation, this regularization term is added to the overall loss function, penalizing the larger weights. As a result, the optimizer updates the weights in such way that it prevents them from growing excessively large which promotes a more balances and stable training process. During gradient descent, larger weights tend to cause larger gradient descent potentially leading to unstable oscillations and convergence in the loss function. The L2 parameter penalizes the weights and naturally dampens the oscillations by shrinking the weight values toward zero in a controlled manner, resulting in smoother convergence process.

2.4.3 Label Smoothing

In medical image classification tasks involving MRI images, model becomes overconfident in their predictions. The network tend to assign very high probabilities (close to 1.0) to one class and extremely low probabilities (close to 0) to all others. This can lead to poor generalization, overfitting and poor convergence rate while dealing with new and unseen data. In order to address this problem, a regularization technique known as label smoothing is employed. In our work, the label smoothing technique was applied at the classification phase to enhance generalization and improve accuracy as well as stabilize the convergence rate.

Label smoothing is a technique where the target class labels are modified during training. Normally , when the label are one-hot encoded, each training label assigns a probability of 1 to the correct label and 0 to the incorrect labels. For example, in the three class classification, the true label for a glioma tumor image would be represented as [1,0,0]. However, with label smoothing the target distribution is "softened" so that the model no longer sees the true class as absolutely certain. The smoothed label will look like [0.95,0.025,0.025] instead. These values depend on the chosen smoothing factor. Mathematically the label smoothing is calculated using the following formula,

$$y_{smooth} = (1 - \alpha) \times y_{true} + \frac{\alpha}{K}$$

Here,

ytrue is the one-hot encoded vector

α is the smoothing factor which is generally between 0 to 1

K is the number of classes which is 3 in this case

In this work, we have chosen a value of 0.05 as the smoothing factor. The figshare brain tumor dataset includes three tumor types : glioma, meningioma and pituitary. The features of these tumors can often overlap in MRI images as intensity distribution, shapes or locations may appear similar in certain cases. For example, a small meningioma near the brain's base might visually resemble a pituitary tumor in some slices. If the model is trained using hard labels or strict one-hot encoding method, then it can become overconfident and memorize some specific patterns instead of learning generalizable representations. Label smoothing solves counters this problem by introducing controlled uncertainty in the training labels. Another advantage of label smoothing is that it helps improve model generalization and reduces overfitting. By softening the labels, the model no longer treats every incorrect class prediction as completely wrong; instead, it penalizes the incorrect outputs proportionally. This smoother error gradient leads to more stable training, especially in multi-class problems like this one, where the number of classes is limited but the visual complexity within each class is high.

2.4.4 Cosine Decay Learning Rate Scheduler

The cosine decay learning rate scheduler is an advanced optimization technique used to dynamically adjust the learning rate during the training of a deep learning model. It plays a crucial role in stabilizing the training , achieving faster convergence and preventing overfitting.

Learning rate is a hyperparameter that controls how much the model's parameters are updated in response to the training during backpropagation. These tools reduce or increase adjust the learning rate as training of the model progresses which helps models to converge more efficiently and lead to better solution. These learning rate schedulers work by changing the learning rate of the model over time depending on some predefined rules or metrics. For example, a learning rate scheduler can reduce the rate over time to allow the model to take more defined steps as it nears the convergence. Others might increase the learning at different pointsto help the model escape plateaus in the loss curve. The goal is to maintain stability and speed. If the learning rate is extremely high then the model may not converge properly and fail to reach an optimal solution. Conversely, if the learning rate is not very much, training can become slow and may get stuck in the local minima. The cosine decay scheduler offers a smooth, non-linear way to reduce the learning rate over time which follows the cosine curve, gradually reducing the learning rate over each cycle.

The cosine decay scheduler follows the following formula for calculating learning rate at each step :

$$\eta_t = \eta_0 \times \frac{1}{2}(1 + cos(\frac{\pi t}{T}))$$

η_t= learning rate at step t

η_0= initial learning rate

T = total number of steps

This formula means that the learning rate initiates its maximum value and gradually reduces to its minimum following a cosine curve shape. This approach ensures smooth transitions rather than abrupt changes in learning rate.

Code snippet for using in training :

```
from tensorflow.keras.optimizers.schedules import CosineDecay
from tensorflow.keras.optimizers import Adam

lr_schedule = CosineDecay(
    initial_learning_rate=0.001,
    decay_steps=19200,
    alpha=0.001
)

optimizer = Adam(learning_rate=lr_schedule)
```

Fig 2.4.2 : Code Snippet for Cosine Decay Learning Rate Scheduler

Here, the initial learning rate sets the starting learning rate at 0.001. The decay steps defines how many steps it will take for the learning rate to complete one full cosine cycle of decay. The alpha parameter specifies that the learning rate does not reach zero but stabilizes instead at a small fraction of the initial value, allowing the model to not stop learning and make small updates even at the end portion of the training. The cosine decay scheduler helps to overcome the overfitting and slow convergence problem by starting from a high learning rate and gradually reducing it as the training progresses. This dynamic control allows the model to explore the parameter space efficiently and refine the learned features carefully with training.

2.5 Performance Metrics

Evaluating a deep learning model performance is an integral part which requires appropriate use of different metrics that accurately measure how well the model is performing. Since our work involves both semantic segmentation and multi class classification different evaluation metrics have been used catered to each task. The performance of segmentation is task is evaluated primarily based on the dice coefficient, dice loss along with accuracy, precision, recall, F1-score and Jaccard Index while the classification performance is assessed using accuracy , precision , recall, F1-score and Categorical Cross-Entropy Loss function.

2.5.1 Dice Coefficient and Dice Loss

Dice Coefficient is a metric used to evaluate the similarity between the predicted mask and ground truth mask. In medical image segmentation, the dice coefficient informs how closely the predicted mask matches the ground truth mask. The calculation of dice coefficient follows the following formula :

$$Dice\ Coefficient\ =\ \frac{2\ |X \cap Y|}{|X|\ +\ |Y|}$$

Here,

X = set of pixels in the predicted mask

Y = set of pixels in the ground truth mask

$X\ \cap\ Y$ = number of overlapping pixels between the prediction and ground truth

In terms of true positives (TP), false positives (FP) and false negatives (FN) :

$$Dice\ Coefficient\ =\ \frac{2\,TP}{2\,TP + FP + FN}$$

A dice score of 1 indicates perfect overlap between prediction and ground truth and a dice score of 0 means there is no overlap. In our work, dice coefficient measures how accurately the model identifies the tumor boundaries from MRI images comparing the ground truth masks. The reason behind using dice coefficient as the performance metric for segmentation model is that traditional accuracy may not effectively capture the segmentation quality as a vast majority of the image is background information which does not include the tumor region. Dice coefficient directly measures the overlapping of target region ensuring more meaningful evaluation of the model performance.

Dice loss is the loss function used in the semantic segmentation model. It is derived from the dice coefficient and is used to optimize the segmentation models. The goal of the model during training is to maximize the dice coefficient which corresponds to minimizing the dice loss. The calculation of dice loss follows the following formula :

Lower dice loss indicates better segmentation performance. Higher loss function directly penalizes the poor overlap between the predicted and actual tumor masks. In the segmentation branch, the dice loss if used to train the network to accurately predict tumor regions. It minimizes the difference between predicted and actual tumor masks, ensuring that all types of tumor regions are captured effectively. Instead of focusing on pixel-wise difference , the loss function leads the model to learn shape and boundary-wise features by focusing on the overlaps.

2.5.2 Accuracy

Accuracy represents the amount of correctly classified samples from the total number of samples. For the multi class classification section the accuracy measures how often the model correctly predicts the tumor type. Each MRI image is assigned one of the three labels (glioma, meningioma or pituitary) and accuracy measures the percentage of correct predictions. It’s a very simple yet effective strategy to measure the model’s performance and obtain an overall view of the model’s predictive power. However, accuracy can be misleading in case of imbalanced datasets as it may be dominated by the majority class. Therefore, complementing with precision, recall and other performance metrics a fair evaluation can be performed. The following formula is for calculating accuracy :

$$Accuracy\ =\ \frac{TP + TN}{TP + TN + FP + FN}$$

Here, TP is the true positive

TN is the true negative

FP is the false positive and FN is the false negative

2.5.3 Precision

Precision or positive predictive value measures how much of the samples predicted as a certain class are actually correct. Precision indicates the frequency with which the model's prediction that an MRI image belongs to a specific tumor class (such as glioma) is accurate. In order to prevent incorrect diagnoses in medical applications, high precision guarantees that the model seldom mislabels non-glioma situations as gliomas. In case of medical image classification higher precision value reduces the rate of false positives ensuring diagnostic reliability. This value is particularly important for medical image classification because in a clinical setting it si important to understand that where the incorrect tumor identification can mislead treatment. High precision indicates that model’s predictions for each tumor class are reliable and more accurate.

The following formula is used for calculating precision :

$$Precision = \frac{TP}{TP + FP}$$

Here, TP is denoted as the true positives and FP is denoted as the false positives.

2.5.4 Recall

Recall, also referred to as Sensitivity or True Positive Rate, evaluates a model’s capability to correctly identify all instances belonging to a particular class. In this study, recall quantifies how effectively the model detects true tumor cases for each specific category, such as glioma, meningioma, or pituitary tumors. A low recall indicates that the model is missing actual tumor cases, leading to false negatives an especially critical issue in medical diagnosis where undetected tumors can have severe consequences. High recall, therefore, signifies that the model is sensitive enough to capture nearly all true tumor occurrences, ensuring that potential cases are not overlooked. This metric holds great clinical importance, as prioritizing recall minimizes the risk of misdiagnosis and enhances patient safety by ensuring that even subtle or complex tumor patterns are successfully identified. The calculation of recall is done based on the following formula :

$$Recall = \frac{TP}{TP + FN}$$

For medical diagnosis using MRI images higher recall values is considered better for the model evaluation.

2.5.5 F1 Score

The F1-score is a crucial evaluation metric that accumulates precision and recall into one single value using their mean, providing a more balanced and comprehensive assessment of model performance particularly in cases of class imbalance.

In the context of brain tumor classification, the F1-score plays a vital role in evaluating how effectively the model identifies different tumor types (such as glioma, meningioma, and pituitary) while simultaneously minimizing both false positives and false negatives. Since medical imaging datasets like the Figshare brain tumor dataset often exhibit class imbalance where certain tumor types have significantly more samples than others, accuracy alone can be misleading. The F1-score compensates for this by ensuring that a high score can only be achieved if the model maintains both strong precision (fewer false alarms) and strong recall (fewer missed detections).

From a clinical perspective, this balance is especially important. A model that achieves high precision but low recall might overlook many real tumor cases, posing serious diagnostic risks. Conversely, a model with high recall but low precision may over-diagnose tumors, causing unnecessary stress or medical intervention. The F1-score, therefore, acts as a harmonizing measure that captures the model's true diagnostic capability across both dimensions.

In summary, the F1-score provides a more reliable and interpretable metric for evaluating classification models in medical imaging, reflecting not only the correctness of predictions but also the model's ability to detect all relevant tumor cases without excessive false alerts. It is thus considered one of the most meaningful indicators of real-world performance in automated brain tumor diagnosis systems.

The formula used for calculating F1 score is mentioned below :

$$F1\,Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

2.5.6 Loss Function

The categorical cross-entropy loss was used as the loss function for optimizing the multi class classification model. Categorical Cross-Entropy (CCE) is one of the most widely used and theoretically grounded loss functions for multi-class classification tasks. It measures the difference between the true class labels and the probability distribution predicted by the model, effectively quantifying how close or far the model's output is from the correct answer. By penalizing incorrect predictions more heavily when the model is confident but wrong, CCE helps guide the model toward more accurate class probability estimations.

Formula is written below :

$$L = -\sum_{i=1}^{C} y_i log(p_i)$$

Here,

C = number of classes

y_i = true class label which is 1 for correct class and 0 otherwise

p_i = predicted probability for class i

In the context of our multi-class brain tumor classification model, the categorical cross-entropy loss evaluates how well the model differentiates between tumor types such as glioma, meningioma, and pituitary. For example, if an MRI image truly belongs to the "glioma" class but the model predicts a higher probability for "pituitary," the loss value increases significantly. This signals the optimizer to adjust the model's parameters in subsequent iterations to improve its predictions. Through repeated training, minimizing CCE ensures that the predicted probability distributions become sharper and more accurate, ideally peaking near the true class while assigning lower probabilities to incorrect classes.

When combined with label smoothing, categorical cross-entropy becomes even more robust by reducing model overconfidence and preventing overfitting. In medical imaging contexts, this feature is particularly valuable because models that are overly certain about their predictions tend to generalize poorly to unseen or slightly varied data. Label smoothing encourages the model to maintain a calibrated confidence level, resulting in more reliable clinical predictions. From a theoretical perspective, CCE is smooth, differentiable, and highly compatible with gradient-based optimization algorithms, making it ideal for deep neural networks such as InceptionV3 used in this project. Its probabilistic foundation also enables uncertainty estimation—an essential factor in medical decision-making, where confidence in classification outcomes directly affects diagnostic reliability.

In summary, categorical cross-entropy serves as the backbone of our classification training process. It not only enforces accurate tumor-type prediction but also stabilizes the learning process, promotes generalization to new MRI samples, and ensures that the model's confidence aligns closely with its true predictive capability ultimately contributing to a more trustworthy and clinically meaningful diagnostic framework.

2.5.7 Jaccard Index (IoU)

The Jaccard index also known as the Intersection over Union (IoU) measures the ratio of the intersection area between the predicted mask and the ground truth mask to their union area. It provides a good measure for the segmentation accuracy.

The Jaccard Index is used as a complementary metric to the Dice Coefficient to assess segmentation quality. In our brain tumor segmentation, it measures how much of the predicted tumor region truly matches the annotated tumor region. Since it penalizes both false positives and false negatives more strictly than Dice, it provides a more conservative evaluation of segmentation accuracy. This value also provides a more realistic estimate of segmentation quality. Ensures the validation and clinical relevance for real world scenarios.

Formula :

$$Jaccard\ Index\ (IoU) = \frac{TP}{TP + FP + FN}$$

Here, TP is denoted as true positives, FP is denoted as false positives and FN is denoted as false negatives.

Chapter 3

3. Results Analysis

3.1 Semantic Segmentation Result

To measure how the segmentation model performs, the Dice Similarity Coefficient (Dice Score) was used as the primary metric. The Dice Score measures the overlap between the predicted segmentation and the ground truth, providing a robust evaluation of segmentation accuracy. Dice similarity coefficient is a widely used metric in segmentation analysis tasks.

Dice Similarity Coefficient (DSC)

The Dice coefficient is defined as:

$$DSC = \frac{2|X \cap Y|}{|X|+|Y|}$$

where:

- X denotes predicted mask,
- Y denotes ground truth mask,
- $|X \cap Y|$ is the intersection of the predicted and actual segmentation,
- smooth is a constant (1e-15) which is added with a target for numerical stability and avoiding division by zero.

The corresponding loss function that is used for model training was the Dice loss:

$$\textbf{DiceLoss} = \textbf{1} - \textbf{DSC}$$

Additionally, the accuracy of the model was also observed, which measures the proportion of correctly classified pixels in the segmentation masks.

The results from all the segmentation models are summarized below:

Table 3.1.1 : Model Evaluation Summary

Model	Accuracy	Dice Coefficient
U Net	99.14%	71.97
U Net + VGG16	99.32%	70.02
U Net + ResNet50	99.31%	65
Attention U Net	99.26%	76.39
Attention Residual U Net	**99.33%**	**79.9**

U Net :

Training accuracy and validation accuracy curve- U net Model :

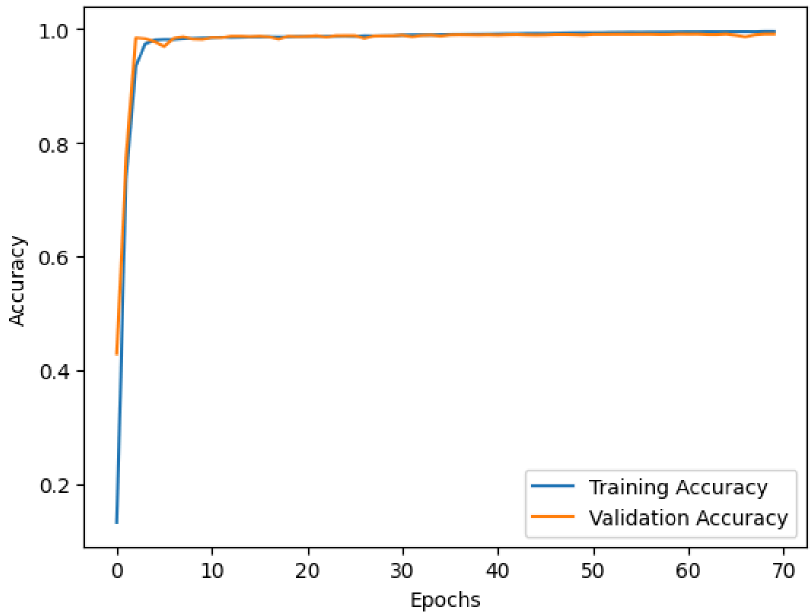


Fig 3.1.1: Training accuracy vs validation accuracy curve for U Net

Training loss and validation loss curve - U Net model :

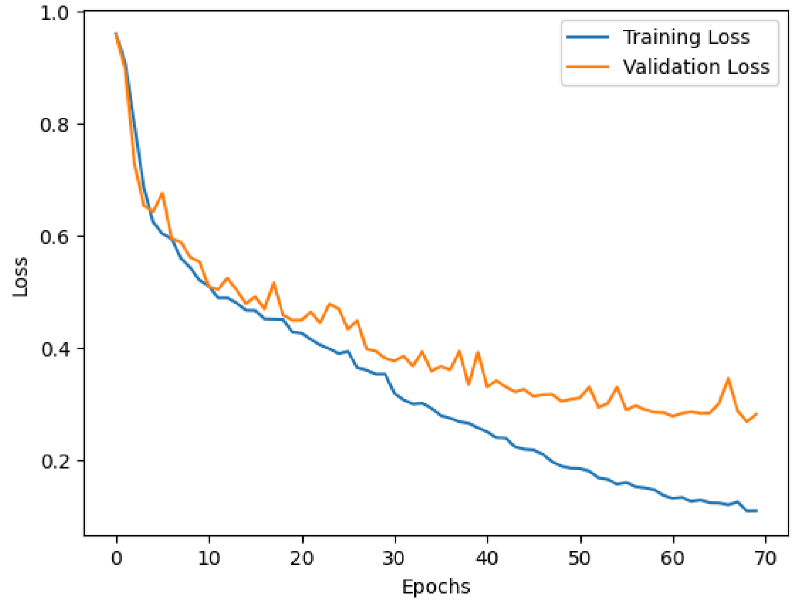


Fig 3.1.2 :Training loss and validation loss curve for U Net model

U net with VGG16 Backbone :

Training accuracy and Validation accuracy curve :

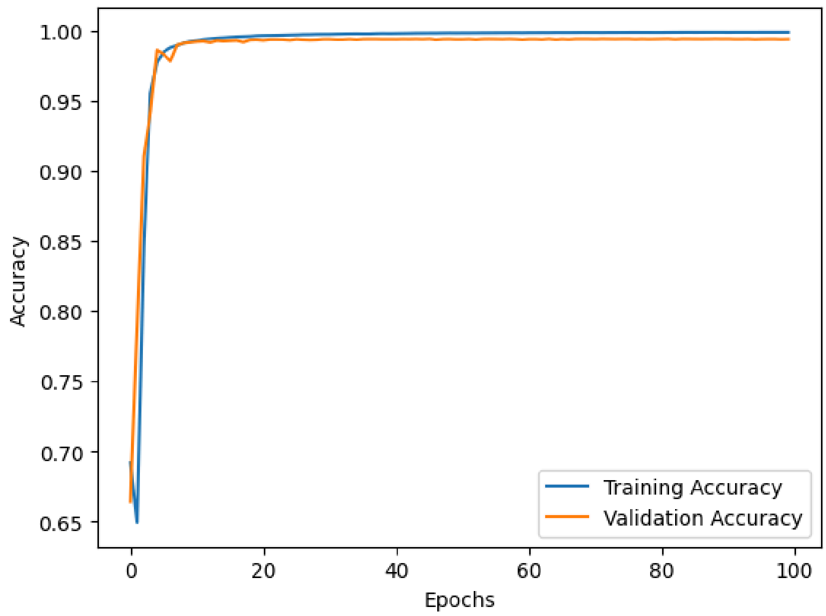


Fig 3.1.3 : Training accuracy and validation accuracy curve for U Net with VGG16

Training loss and Validation loss curve :

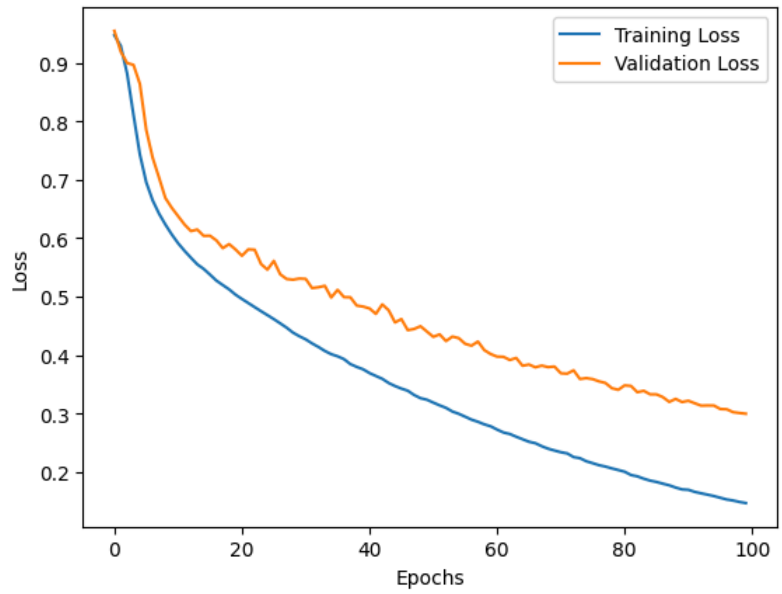


Fig 3.1.4 : Training loss and validation loss curve for U Net with VGG16

U net with ResNet50 Backbone :

Training accuracy and Validation accuracy curve :

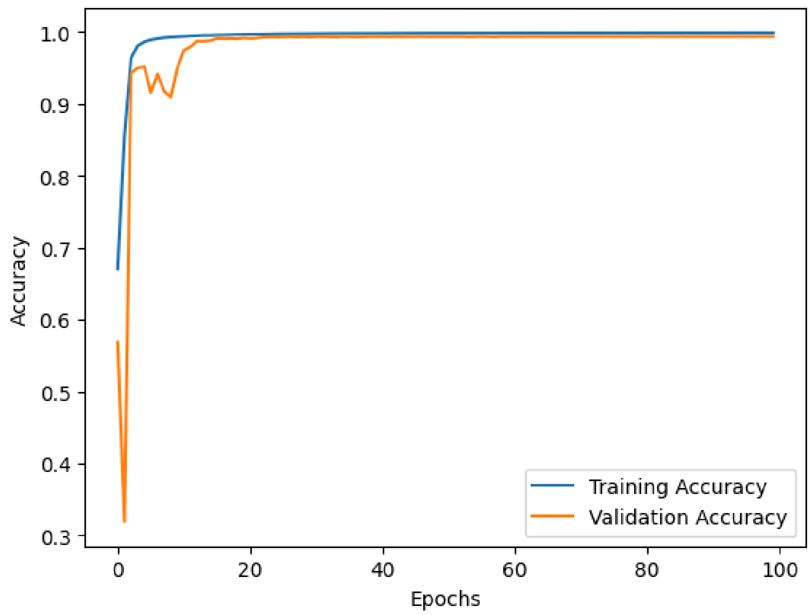


Fig 3.1.5: Training accuracy and validation accuracy curve for U Net with ResNet50

Training loss and Validation loss curve :

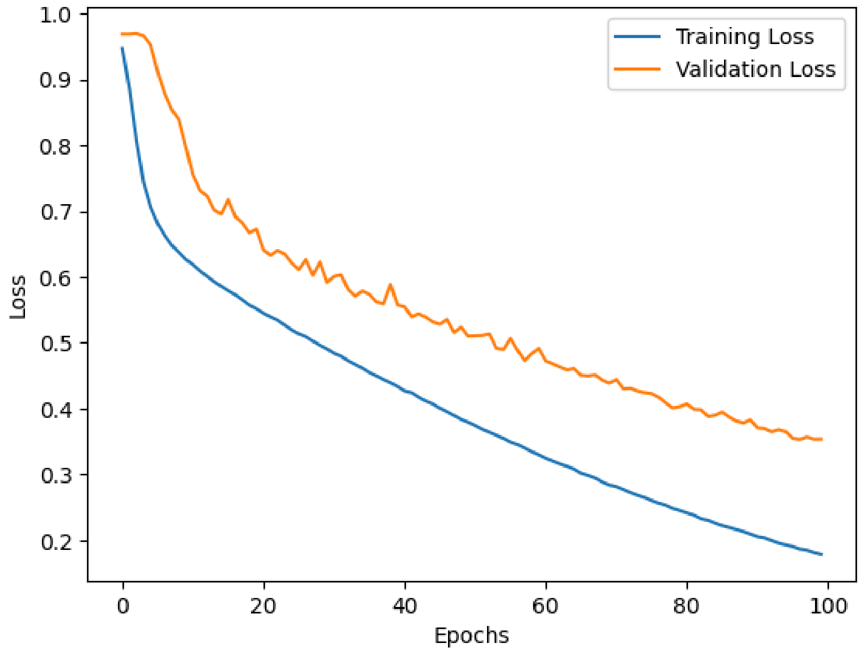


Fig 3.1.6: Training loss and validation loss curve for U Net with ResNet50

Attention U Net :

Training accuracy and Validation accuracy curve :

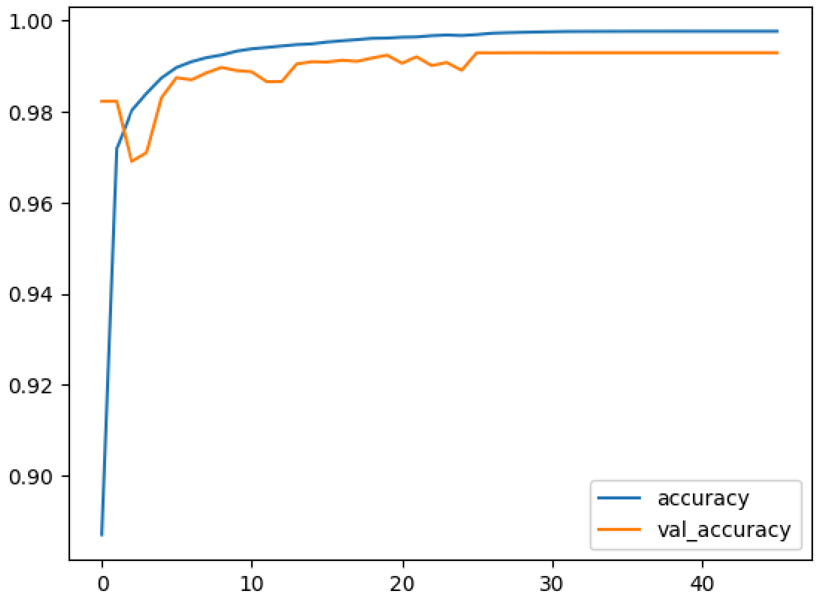


Fig 3.1.7: Training accuracy and validation accuracy for Attention U Net

Training loss and Validation loss curve :

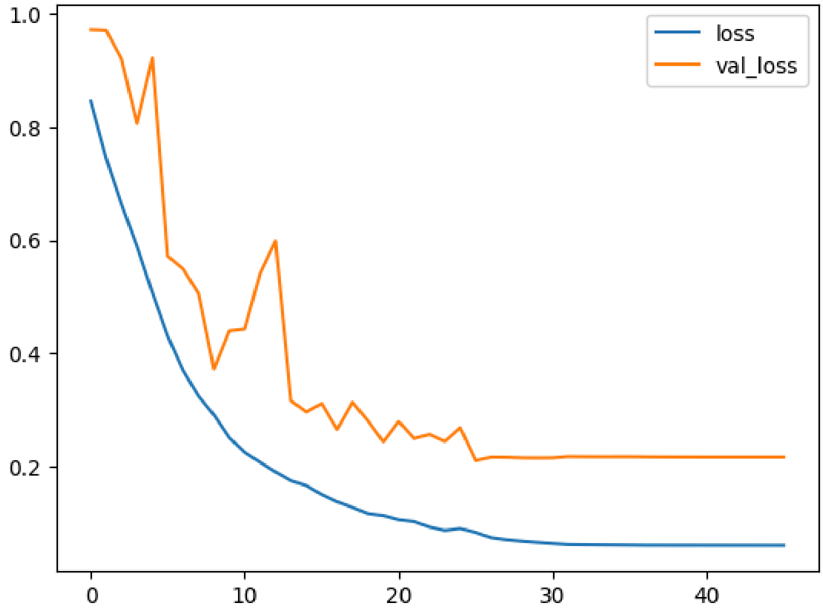


Fig 3.1.8: Training loss and validation loss curve for Attention U Net

Attention Gate Residual U Net :

Training accuracy and Validation accuracy curve :

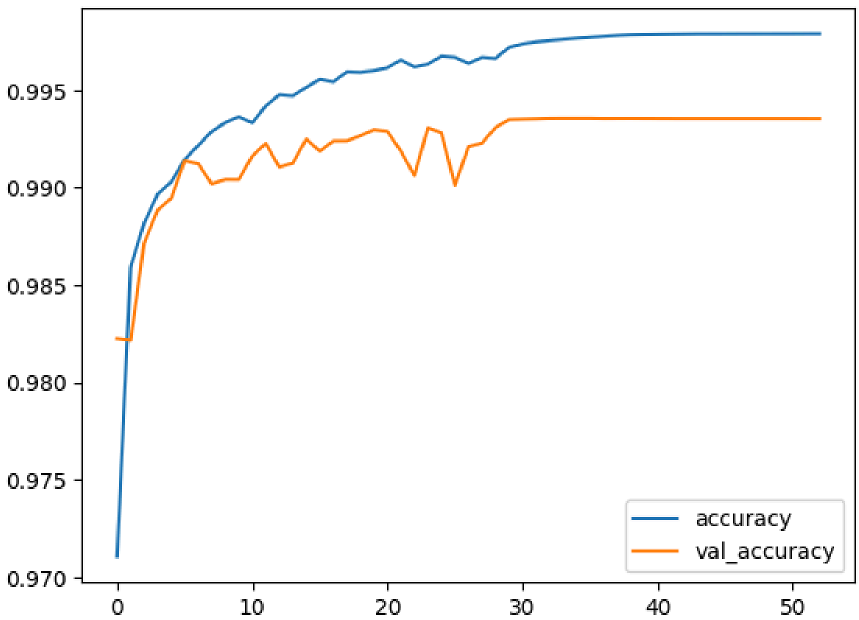


Fig 3.1.9: Training accuracy and validation accuracy curve for Attention Residual U Net

Training loss and Validation loss curve :

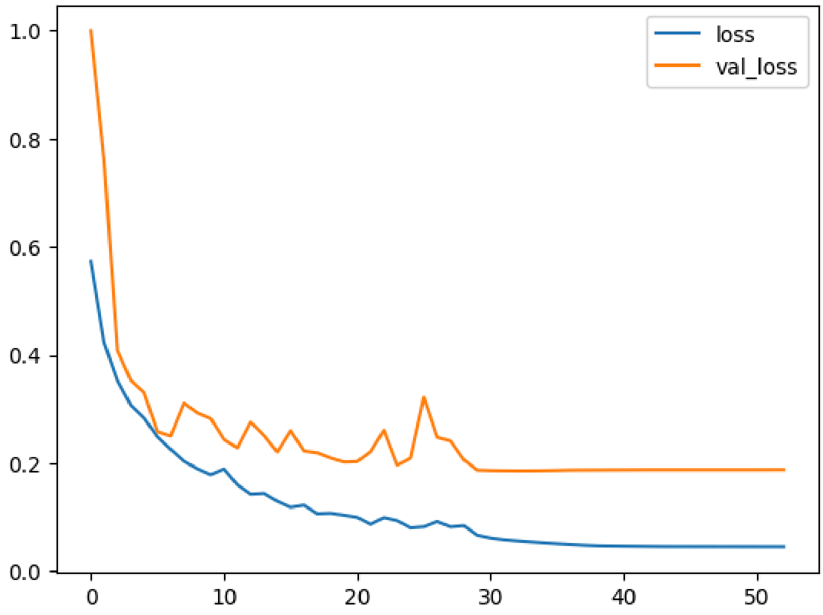


Fig 3.1.10 : Training loss and validation loss curve for Attention Residual U Net

Prediction :

Prediction for U Net Model :

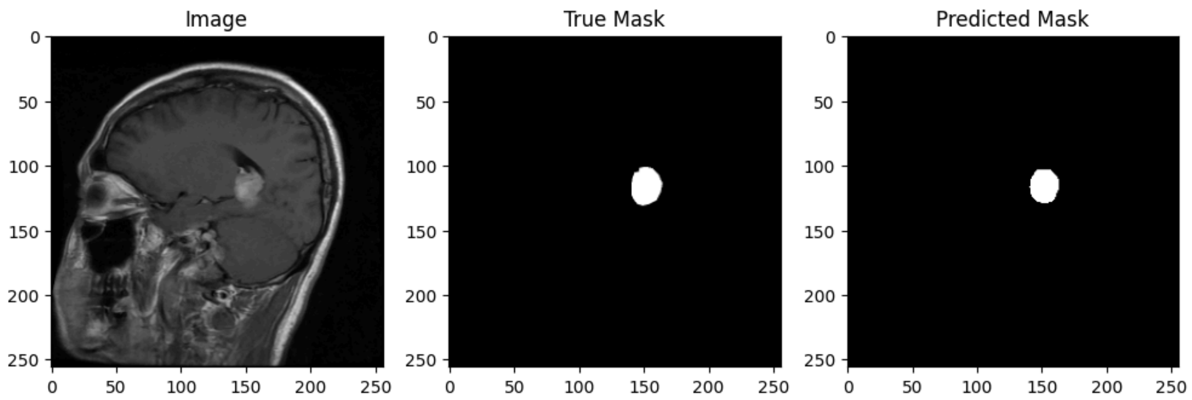


Fig 3.1.11: Prediction for U Net

Prediction for U Net with VGG16 Backbone :

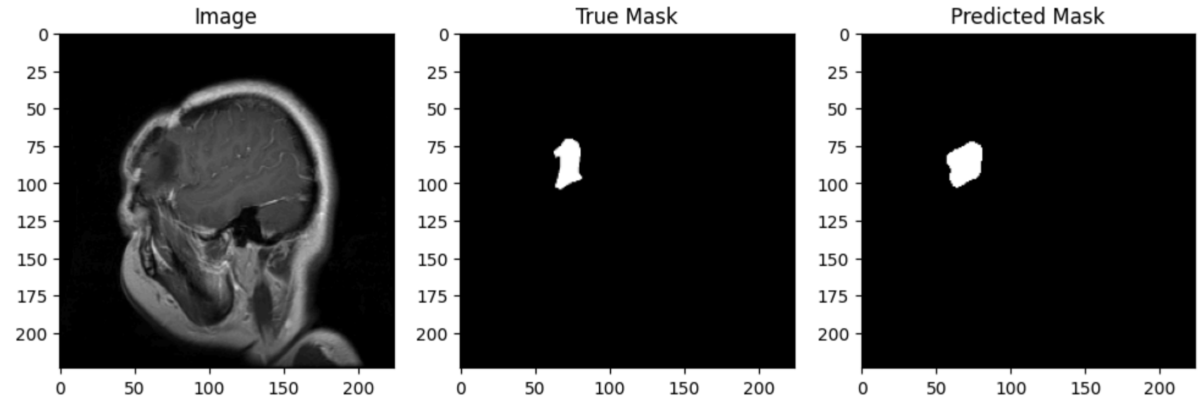


Fig 3.1.12: Prediction for U Net with VGG16

Prediction for U Net with ResNet50 Backbone :

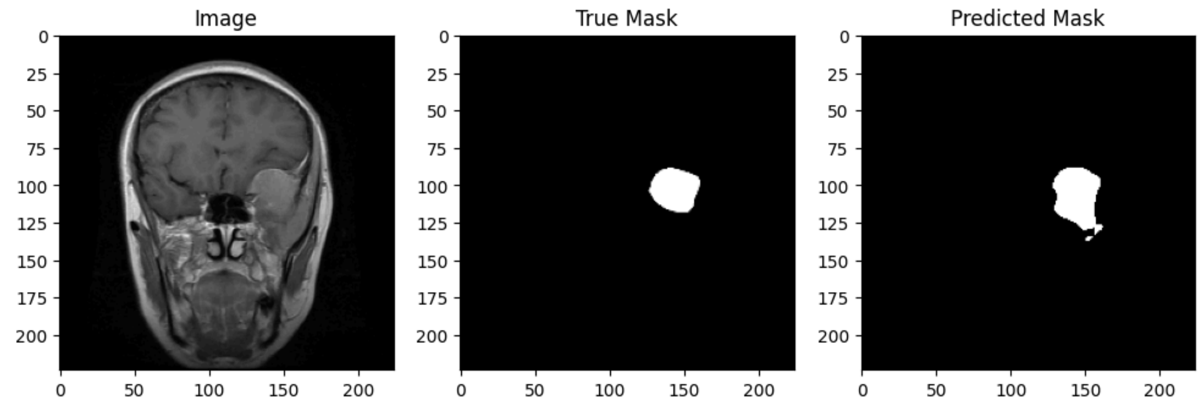


Fig 3.1.13: Prediction for U Net with ResNet50

Prediction for Attention U Net :

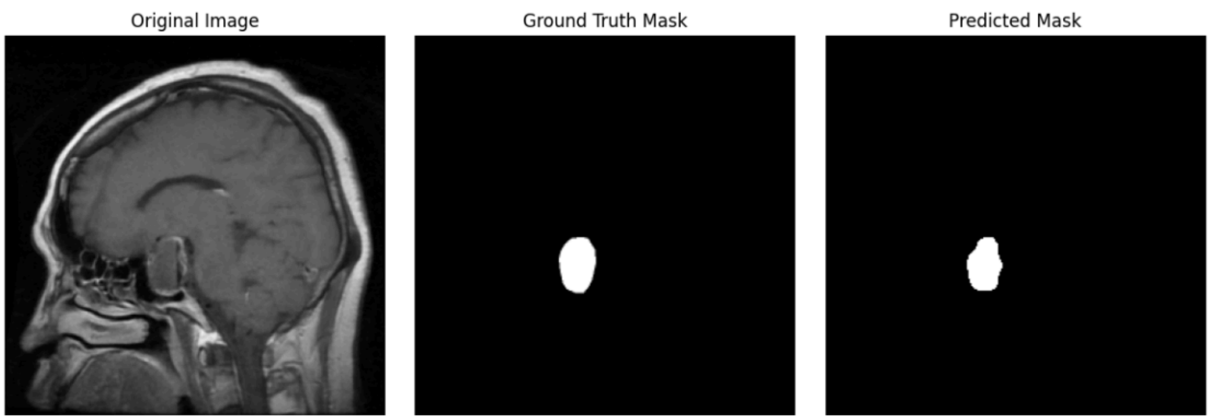


Fig 3.1.14 : Prediction for Attention U Net

Prediction for Residual Attention U Net :

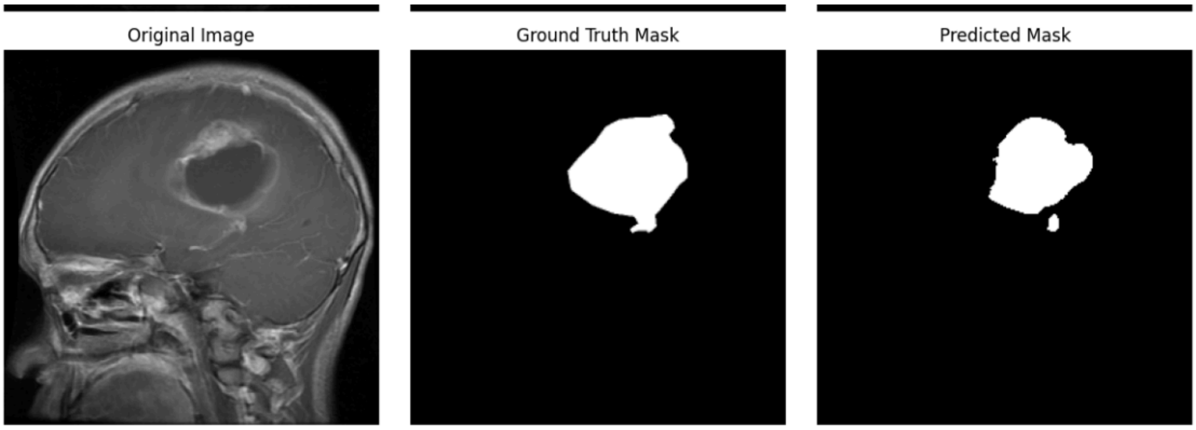


Fig 3.1.15: Prediction for Attention Residual U Net

The final Attention Gate Residual U Net model has yielded the highest dice score among all the brain tumor segmentation models. Applying an attention gate along the native U net architecture has significantly improved the model performance. Hence, the predictions generated from this model was further preprocessed for the multi class classification model.

Table 3.1.2 : Attention Residual U Net model Performance Summary

Metric	**Value**
F1 Score	0.74433
Jaccard Index (IoU)	0.66369
Recall	0.74380
Precision	0.79460
Test Accuracy	0.9933
Dice Coefficient	0.80

3.2 Multi Class Classification Result

After obtaining the predicted tumor masks from the Attention Residual U-Net segmentation model, the binary masks were overlapped with the original MRI images to form composite two-channel tensors, later expanded to three channels for classification. These enhanced images were used to train and test the InceptionV3 classifier within the proposed multi-task framework. The model was trained to classify tumors into three major categories. The performance metrics obtained for our proposed Multi-task learning framework:

Table 3.2.1 : InceptionV3 Classifier Performance Summary

Accuracy	93.23%
Precision	0.92
Recall	0.91
F1 Score	0.92

Training vs Validation accuracy of the InceptionV3 classifier:

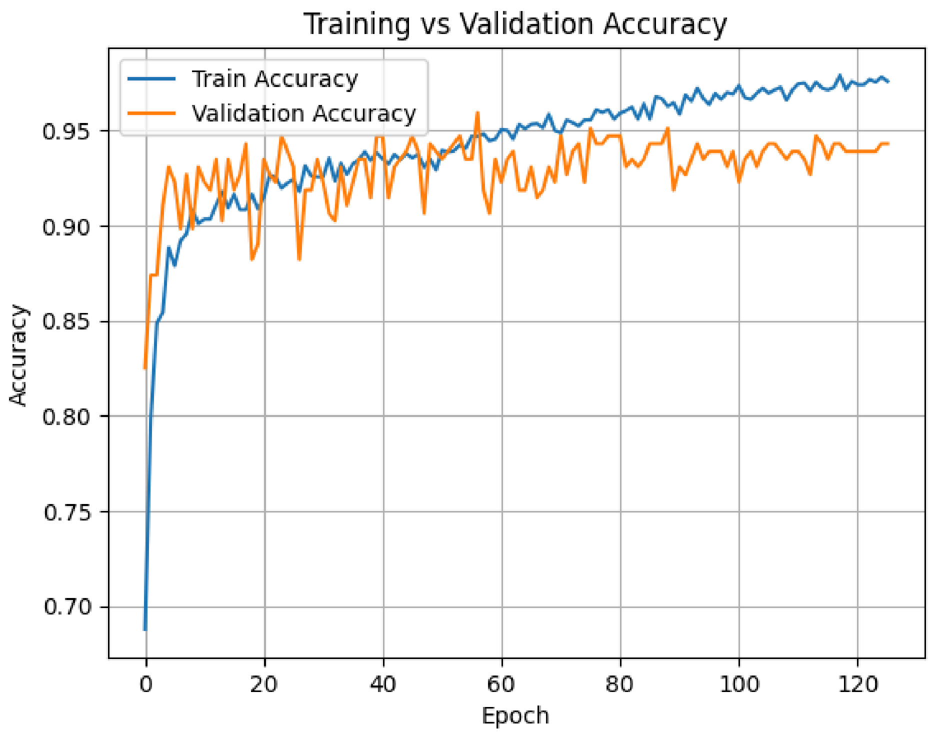


Fig 3.2.1: Training vs Validation Accuracy Curve for Classifier

Training vs Validation Loss of the InceptionV3 classifier:

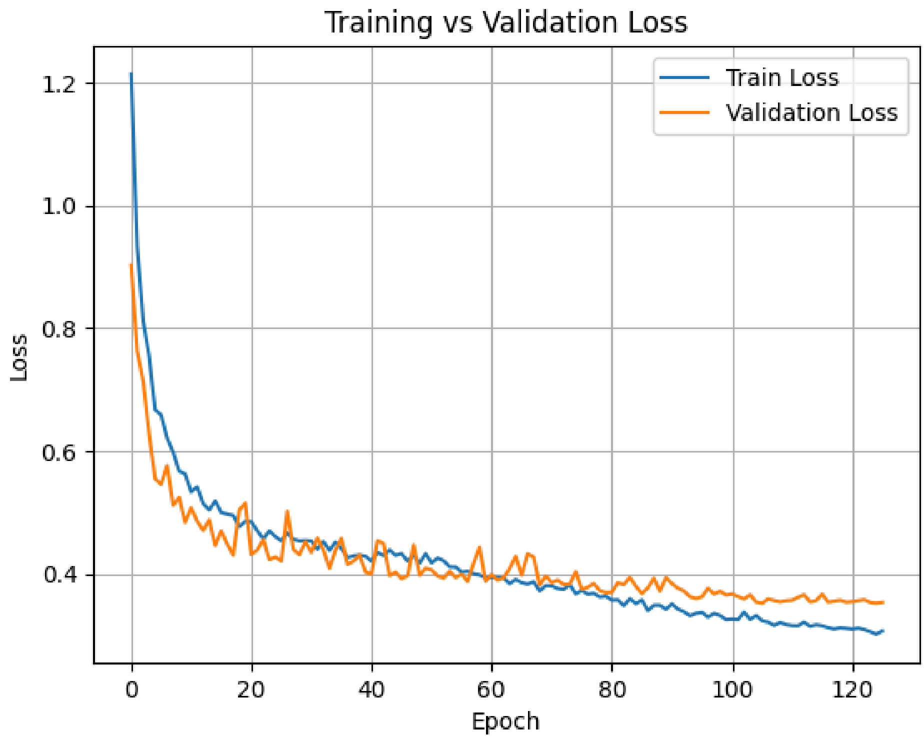


Fig 3.2.2: Training vs Validation Loss Curve for Classifier

During the training phase, the model demonstrated stable learning behavior. The training vs. validation accuracy curve (Figure 3.2.1) shows a rapid increase in accuracy during the first few layers, followed by a gradual stabilization as the network converges. Training accuracy continued to improve steadily, reaching approximately 96–97% by the final epochs, while validation accuracy stabilized around 93–94%, indicating effective generalization without significant overfitting.

Similarly, the loss curve (Figure 3.2.2) exhibited a consistent downward trend for both training and validation sets. The training loss decreased sharply in the first few epochs and continued to decline at a slower rate as optimization progressed, whereas the validation loss plateaued with minor oscillations, signifying a well-regularized model.

Overall, the performance curves affirm that the proposed InceptionV3-based classification network achieved both high accuracy and stable convergence.

3.2.1 Classwise Performance Metrics

The multi-class classification prediction performance for each tumor type is summarized in the following table. The glioma and pituitary classes achieved the highest precision and recall scores, while the meningioma class showed slightly lower accuracy due to class imbalance. Nevertheless, the F1-scores across all three classes remained consistently high, confirming the model's balanced performance.

Table 3.2.2 : Classwise Performance Metrics

Class	**Accuracy**	**Precision**	**Recall**
Meningioma	82%	0.89	0.82
Glioma	95%	0.94	0.96
Pituitiary	95%	0.93	0.96

3.2.2 Comparison with CNN Classifiers

To evaluate the advantage of our proposed framework, we compared it with several state-of-the-art CNN architectures trained on the same dataset without segmentation guidance. The comparison models included VGG16, VGG19, InceptionV3, EfficientNetV2S and ConvNeXtTiny.

Table 3.2.3 :Comparison of Proposed Framework with CNN Classifiers

Model	**Accuracy**	**Precision**	**Recall**	**F1 score**
InceptionV3	0.8760	0.83	0.39	0.84
VGG16	0.93	0.90	0.22	0.87
VGG19	0.9315	0.90	0.23	0.87
EfficientNetV2S	0.9103	0.86	0.31	0.86
ConvNextTiny	0.9103	0.88	0.27	0.89

Multitask Learning Framework	**0.9323**	**0.92**	**0.91**	**0.92**

The proposed multi-task InceptionV3 classifier consistently outperformed other CNN architectures in terms of accuracy, precision, recall and F1-score. The integration of the segmentation mask provides spatial attention, enabling the model to focus on the region of interest (ROI) and thus reduce false predictions caused by non-tumorous regions.

3.2.3 Confusion Matrix

The confusion matrix for the test set is presented below. Each row of the matrix shows the true class, while every column depicts the predicted class. The diagonal entries indicate correct classifications and off-diagonal entries denote misclassifications.

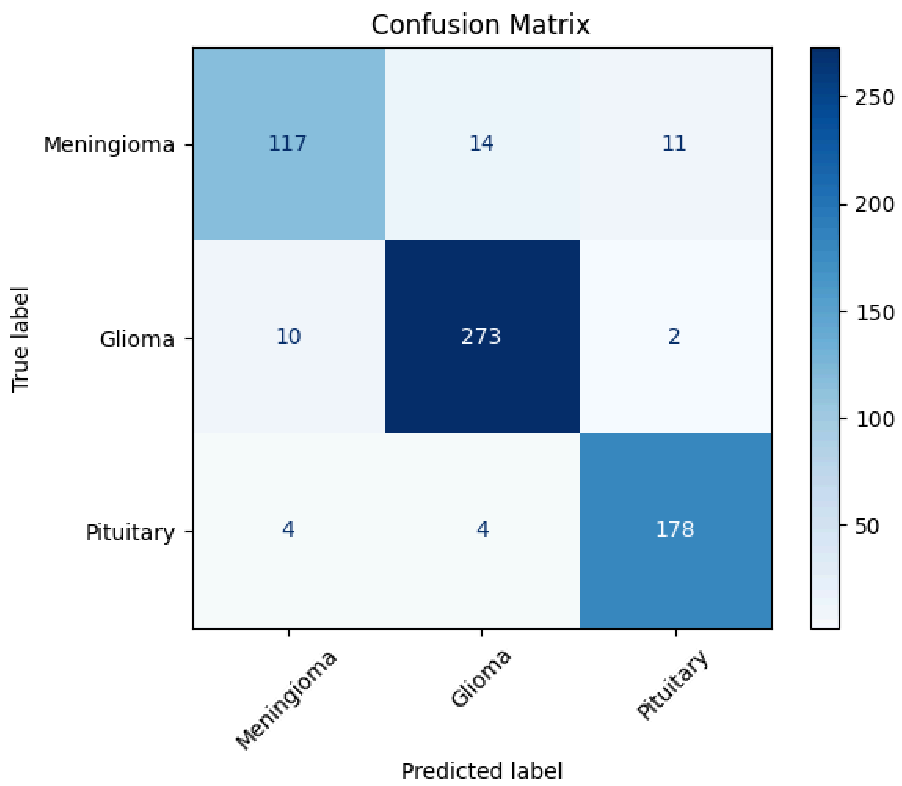


Fig 3.2.3: Confusion Matrix

Table 3.2.4: Analysis from Confusion Matrix

True Class	Predicted Meningioma	Predicted Glioma	Predicted Glioma
Meningioma	117	14	11
Glioma	10	273	2
Pituitiary	4	4	178

From the confusion matrix, it can be observed that the glioma and pituitary classes exhibit very high true-positive counts with minimal cross-class confusion. The type meningioma shows a little higher misclassification rate, mainly being confused with glioma, likely due to overlapping visual characteristics in T1-weighted MRI images.

Sample outputs from the test dataset are shown in the following, where the predicted and true tumor types are displayed side by side with their corresponding MRI images.

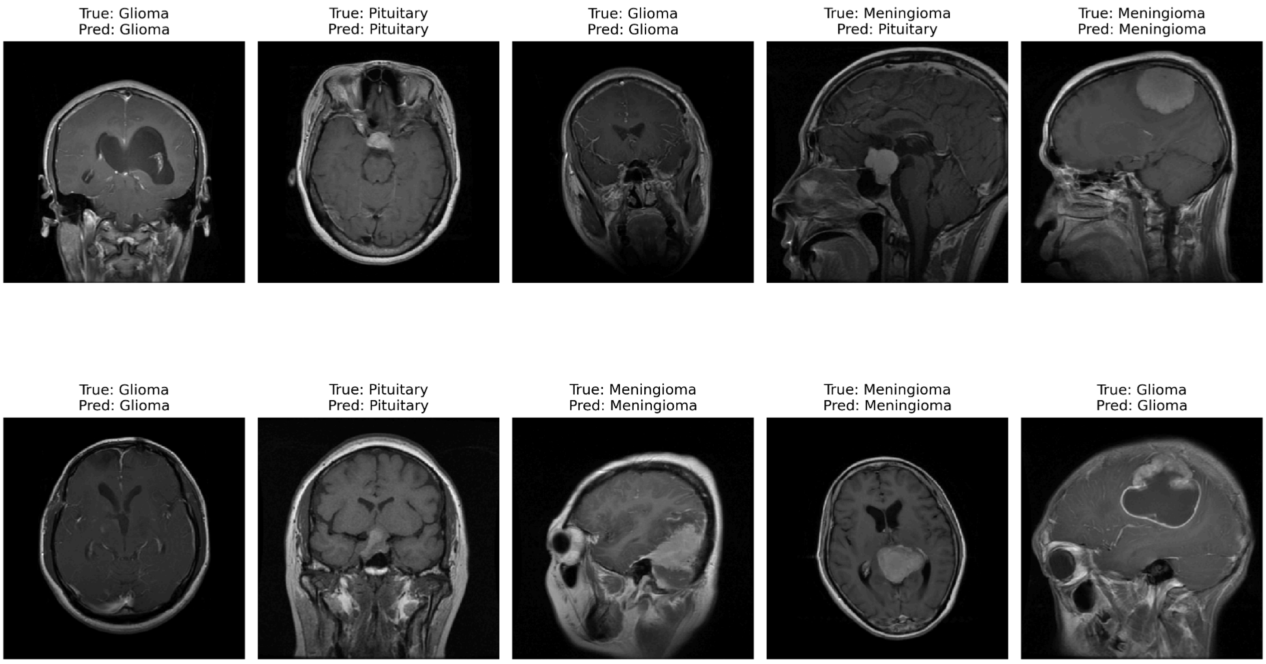


Fig 3.2.4.: Final Predictions

Chapter 4

4. Demonstration of Outcome Based Education (OBE)

4.1 Introduction

In recent years, the industry has witnessed rapid advancements in the fields of computer vision and deep learning, leading to transformative applications in biomedical image processing. The integration of artificial intelligence (AI) into healthcare has enabled unprecedented capabilities in diagnostic imaging, automating tasks that once relied heavily on manual interpretation by medical professionals. From detecting anomalies in X-rays and CT scans to segmenting tumors in MRI images, deep-learning models have significantly enhanced the speed, accuracy, and reliability of medical diagnoses.

Against this backdrop of technological progress, our thesis project focuses on addressing a critical challenge in the domain of neuro-oncology: the accurate and efficient detection and classification of brain tumors from MRI scans. Leveraging state-of-the-art convolutional neural networks and multi-task learning frameworks, our work seeks to automate the processes of tumor segmentation and classification within a unified model. This approach not only aims to improve diagnostic precision but also to reduce the workload on radiologists, expedite clinical decision-making, and contribute to equitable access to healthcare in regions lacking specialized expertise.

This chapter demonstrates how our project aligns with the principles of Outcome-Based Education (OBE) by mapping the course outcomes, program outcomes, and knowledge profiles addressed through the research process. By connecting theoretical knowledge, practical implementation, ethical considerations, and societal impacts, this project embodies the multidisciplinary, complex problem-solving approach central to engineering education and practice.

4.2 Course Outcomes (COs) Addressed

The following table shows the COs addressed in EEE 4700 for our Project and Thesis.

COs	CO Statement	POs	Put Tick (√) EEE 4800
CO1	Apply knowledge of electrical and electronic engineering to the solution of complex engineering problems.	PO1	√
CO2	Identify a contemporary real life problem related to electrical and electronic engineering by reviewing and analyzing existing research works.	PO2	√
CO3	Determine functional requirements of the problem considering feasibility and efficiency through analysis and synthesis of information.	PO4	√
CO4	Select a suitable solution and determine its method considering professional ethics, codes and standards.	PO8	√
CO5	Adopt modern engineering resources and tools for the solution of the problem.	PO5	√
CO6	Prepare management plan and budgetary implications for the solution of the problem.	PO11	√
CO7	Analyze the impact of the proposed solution on health, safety, culture and society.	PO6	√
CO8	Analyze the impact of the proposed solution on environment and sustainability.	PO7	√
CO9	Develop a viable solution considering health, safety, cultural, societal and environmental aspects.	PO3	√
CO10	Work effectively as an individual and as a team member for the accomplishment of the solution.	PO9	√
CO11	Prepare various technical reports, design documentation, and deliver effective presentations for demonstration of the solution.	PO10	√
CO12	Recognize the need for continuing education and participation in professional societies and meetings.	PO12	√

COs	Explanation/Justification
CO1	This project identifies the contemporary issue of brain tumor detection and diagnosis by reviewing existing research on MRI-based methods and analyzing the limitations of manual segmentation and classification, highlighting the need for an automated deep learning solution.
CO2	The proposed architecture determines the functional requirements by analyzing the feasibility of using multi-task deep learning to perform both tumor classification and segmentation efficiently. The framework is designed to optimize accuracy while ensuring it can process multimodal MRI data within practical computational constraints.
CO3	This project selects a deep learning-based multi-task framework as the solution, adhering to professional ethics by utilizing publicly available datasets (BraTS2020) and ensuring transparency in model design. It follows industry standards in medical imaging and AI, ensuring the solution is ethical, reproducible, and compliant with established codes for handling medical data.
CO4	Our proposal adopts modern engineering resources and tools, including convolutional neural network architectures like U-Net, VGG16, and ResNet50, along with deep learning frameworks to effectively develop and implement the multi-task learning framework for brain tumor segmentation and classification.

CO5	Our research progression will be done through the proper framework. From software implementation to hardware implementation proper management needs to be followed for the succession of our research work. For the hardware implementation of our proposal, we need several devices where pricing is involved.
CO6	The proposed solution enhances health outcomes through timely and accurate diagnoses, promotes equitable access to care, and fosters a culture of technological integration in healthcare, thereby improving overall societal well-being.
CO7	The proposed solution contributes to sustainability by reducing the need for invasive diagnostic procedures and promoting efficient resource use in healthcare, ultimately minimizing environmental impact.
CO8	This project devises a feasible approach by prioritizing health and safety concerns within the diagnostic procedure, accounting for societal and cultural factors to promote equitable access to healthcare, and mitigating environmental impact through sustainable resource utilization.
CO9	This project emphasizes collaboration among teammates by encouraging communication and coordination among members while also promoting individual responsibility to ensure the successful development and implementation of the proposed solution.
CO10	The project involves the preparation of comprehensive technical reports and design documentation, along with presentations, to clearly communicate the solution's methodology, results, and implications to stakeholders and the broader community.
CO11	This project recognizes the significance of continuous professional development and fosters engagement in professional communities and conferences to stay updated of the latest innovations in deep learning and medical imaging technologies.

4.3 Aspects of Program Outcomes (POs) Addressed

The following table shows the aspects addressed for certain Program Outcomes (POs) addressed in EEE 4700 for Project and Thesis.

	Statement	**Different Aspects**	**Put Tick (√)**
PO3	**Design/development of solutions**: Design solutions for complex electrical and electronic engineering problems and design systems, components or processes that meet specified needs with appropriate consideration for public health and safety, cultural, societal, and environmental considerations.	Public health	√
		Safety	√
		Cultural	
		Societal	√
		Environmental	
PO4	**Investigation:** Conduct investigations of complex electrical and electronic engineering problems using research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of information to provide valid conclusions.	Design of experiments	√
		Analysis and interpretation of data	√
		Synthesis of information	√
PO6	**The engineer and society:** Apply reasoning informed by contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to	Societal	√
		Health	√
		Safety	√

	professional engineering practice and solutions to complex electrical and electronic engineering problems.	Legal	
		Cultural	
PO7	**Environment and sustainability:** Understand and evaluate the sustainability and impact of professional engineering work in the solution of complex electrical and electronic engineering problems in societal and environmental contexts.	Societal	√
		Environmental	√
PO8	**Ethics:** Apply ethical principles embedded with religious values, professional ethics and responsibilities, and norms of electrical and electronic engineering practice.	Religious values	
		Professional ethics and responsibilities	√
		Norms	√
PO10	**Communication:** Communicate effectively on complex engineering activities with the engineering community and with society at large, such as being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.	Comprehend and write effective reports	√
		Design documentation	√
		Make effective presentations	√
		Give and receive clear instructions	√
PO11	**Project management and finance:** Demonstrate knowledge and understanding of engineering management principles and economic decision-making and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.	Engineering management principles	√
		Economic decision-making	√
		Manage projects	√
		Multidisciplinary environments	√

4.4 Knowledge Profiles (K3 – K8) Addressed

The following table shows the Knowledge Profiles (K3 – K8) addressed in EEE 4700 for Project and Thesis:

K	**Knowledge Profile (Attribute)**	**Put Tick (√)**
K3	A systematic, theory-based formulation of engineering fundamentals required in the engineering discipline	√
K4	Engineering specialist knowledge provides theoretical frameworks and bodies of knowledge for the accepted practice areas in the engineering discipline; much is at the forefront of the discipline	√
K5	Knowledge that supports engineering design in a practice area	√
K6	Knowledge of engineering practice (technology) in the practice areas in the engineering discipline	√
K7	Comprehension of the role of engineering in society and identified issues in engineering practice in the discipline: ethics and the engineer's professional responsibility to public safety; the impacts of engineering activity; economic, social, cultural, environmental and sustainability	√
K8	Engagement with selected knowledge in the research literature of the discipline	√

K	Explanation/Justification
K3	The project applies fundamental deep learning and convolutional neural network (CNN) principles, crucial in engineering for designing automated classification and segmentation systems.
K4	By leveraging state-of-the-art architectures like U-Net, VGG16, and ResNet, the project operates at the forefront of engineering knowledge for medical image analysis.
K5	The project integrates theoretical knowledge with practical design, aiming to develop a multi-task deep learning model for brain tumor classification and segmentation.
K6	The project involves applying advanced technological tools, such as deep learning frameworks and MRI imaging, in a highly specialized field of medical image processing.
K7	The project addresses societal and ethical concerns by contributing to non-invasive, efficient diagnosis of life-threatening brain diseases, highlighting the role of engineering in healthcare and its impact on sustainability and safety.
K8	The project engages with current research by using datasets like BraTS2020 and builds on existing knowledge in the field of neural networks and medical image analysis, pushing forward the boundaries of current techniques.

4.5 Use of Complex Engineering Problems

- **Interdisciplinary Integration**: The project combines medical knowledge (brain tumor characteristics from MRI scans) with advanced engineering techniques (deep learning and convolutional neural networks), addressing the complexity of multi-modal data processing.
- **Multi-task Learning**: By jointly solving tumor classification and segmentation in one framework, the project tackles multiple objectives simultaneously, optimizing resources and improving efficiency, which addresses the complexity of handling distinct yet related tasks.
- **Advanced Model Architectures**: The use of U-Net for segmentation and models like VGG16, ResNet for classification demonstrates the application of sophisticated neural networks that are designed to handle intricate patterns in MRI images, providing precise and reliable outputs.
- **Data Complexity and Preprocessing**: The project deals with complex MRI data from multiple modalities (T1, T2, FLAIR, etc.), requiring advanced preprocessing and augmentation techniques to ensure the model can generalize across varying data types and tumor characteristics.
- **Evaluation Metrics**: To ensure the model's robustness, performance is evaluated using rigorous metrics like Dice similarity coefficient, IoU, accuracy, and F1 score, which are designed to handle the inherent complexities of medical image segmentation and classification.
- **Scalability and Efficiency**: The project addresses the challenge of manual, time-consuming processes in medical diagnosis by developing a scalable and efficient automated solution, enabling faster and more accurate diagnosis in clinical settings.

4.6 Socio-Cultural, Environmental, And Ethical Impact

Brain tumors are amongst the most life-threatening diseases in all age groups. According to a statistical report, there have been 25,400 new brain tumor cases in 2024. The treatment procedure for brain tumors is very time-consuming due to the lack of automation in the diagnostic process. This project aims to mitigate this challenge by implementing a multi-task learning framework based on a convolution neural network for segmentation and classification of brain tumors from MRI images. This task has always been done manually by radiologists, introducing automation in the diagnostic procedure of brain tumor detection would lead to a positive social impact. Accurate and timely diagnosis would be able to save hundreds of lives and increase longevity in many more. The research could help reduce the disparity in access to high-quality medical care, particularly in areas where skilled radiologists might be scarce. The ethical consideration has to be taken into account while working with medical data in order to protect patients' personal information.

4.7 Attributes of Ranges of Complex Engineering Problem Solving (P1 – P7) Addressed

The following table shows the attributes of ranges of Complex Engineering Problem Solving (P1 – P7) addressed in EEE 4700 for Project and Thesis.

P	**Range of Complex Engineering Problem Solving**	**Put Tick (√)**
Attribute	Complex Engineering Problems have characteristics P1 and some or all of P2 to P7:	
Depth of knowledge required	**P1:** Cannot be resolved without in-depth engineering knowledge at the level of one or more of K3, K4, K5, K6, or K8 which allows a fundamentals-based, first principles analytical approach	√
Range of conflicting requirements	**P2:** Involve wide-ranging or conflicting technical, engineering, and other issues	√
Depth of analysis required	**P3:** Have no obvious solution and require abstract thinking, and originality in analysis to formulate suitable models	√
Familiarity of issues	**P4:** Involve infrequently encountered issues	√
Extent of applicable codes	**P5:** Are outside problems encompassed by standards and codes of practice for professional engineering	√
Extent of stakeholder involvement and conflicting requirements	**P6:** Involve diverse groups of stakeholders with widely varying needs	√
Interdependence	**P7:** Are high level problems including many component parts or sub-problems	√

The following table explains or justifies how the attributes of ranges of Complex Engineering Problem Solving (P1 – P7) have been addressed in EEE 4700/4800 (Project and Thesis).

P	Explanation/Justification
P1	The project requires in-depth knowledge of deep learning (K4), CNN architectures, and medical imaging principles (K5), utilizing first principles of neural networks and image processing to develop an automated system for tumor classification and segmentation.
P2	The project must balance the accuracy and efficiency of both segmentation and classification tasks, optimize computational resources, and account for variations in MRI modalities and tumor types, which can introduce technical challenges and conflicts between precision and performance.
P3	No clear, established solution exists for a fully automated, multi-task model that performs both segmentation and classification simultaneously. The project requires abstract thinking and innovation to design a shared encoder and tackle this challenge efficiently through novel neural network structures.
P4	Brain tumor segmentation and multi-class classification are specialized and infrequently encountered issues, particularly in the development of a unified model for both tasks. This introduces unfamiliar, complex scenarios in medical image analysis.
P5	The problem lies outside typical engineering standards and codes of practice, as it focuses on cutting-edge applications of AI in medical diagnostics. There are no universally accepted frameworks or codes specifically for such automated medical imaging tasks.
P6	The project must address the needs of multiple stakeholders, including medical professionals (neuro-radiologists), researchers, and engineers. Each group has differing requirements, such as accuracy for diagnosis, scalability, and usability in clinical settings, which must be harmonized.
P7	The project is highly interdependent, involving several sub-problems, such as accurate MRI data preprocessing, effective feature extraction from multi-modal images, segmentation of irregular tumor boundaries, and classification of tumor types, all requiring integration into a cohesive, functioning system.

4.8 Attributes of Ranges of Complex Engineering Activities (A1 – A5) Addressed

The following table shows the attributes of ranges of Complex Engineering Activities (A1 – A5) addressed in EEE 4700 for Project and Thesis.

A	**Range of Complex Engineering Activities**	**Put Tick (√)**
Attribute	Complex activities means (engineering) activities or projects that have some or all of the following characteristics:	
Range of resources	**A1:** Involve the use of diverse resources (and for this purpose resources include people, money, equipment, materials, information and technologies)	√
Level of interaction	**A2:** Require resolution of significant problems arising from interactions between wide-ranging or conflicting technical, engineering or other issues	√
Innovation	**A3:** Involve creative use of engineering principles and research-based knowledge in novel ways	√
Consequences for society and the environment	**A4:** Have significant consequences in a range of contexts, characterized by difficulty of prediction and mitigation	√
Familiarity	**A5:** Can extend beyond previous experiences by applying principles-based approaches	√

The following table explains or justifies how the attributes of ranges of Complex Engineering Activities (A1 – A5) have been addressed in EEE 4700/4800 (Project and Thesis).

A	Explanation/Justification
A1	The project requires MRI brain image data from different databases collected from various different sources. It also requires equal collaboration between the medical domain of brain imaging modality as well as the concept and application of artificial neural networks and computer vision to completely reach all the aforementioned goals of the project.
A2	This project addresses a wide range of significant issues ranging from designing an automated process for brain tumor detection to the integration of technologies from the medical and engineering fields. In the development of a multi-task learning framework, the issues of dealing with medical images from a technical aspect have to be resolved.
A3	In this project, models developed based on convolutional neural networks will be used for the detection and segmentation of brain tumors from MRI images. Computer vision-based principles will be used to leverage deep learning networks in order to build the multi-task learning framework.
A4	The proposed architecture of deep learning networks addresses significant issues in biomedical imaging, computer vision as well as artificial intelligence-based deep learning networks. Coordinating the medical field along with the technical aspect of computer vision and neural networks is a time-consuming and complex task that will be mitigated by this project.
A5	The proposed work is based on the learnings and state-of-the-art models of previous research work on biomedical imaging as well as artificial neural network-based models. The proposed model builds upon previous research works and takes inspiration from researchers pursuing works on biomedical imaging modalities and the improvement of different methodologies of visualizing brain tumor images.

5. Conclusion

The research proposed and implemented a multi-task deep learning framework for the automatic segmentation and multi-class classification of brain tumors from MRI scans.

The proposed framework was designed in two major stages: (1) semantic segmentation and (2) classification. In the segmentation stage, an Attention Residual U-Net model was employed to generate precise binary tumor masks. The incorporation of residual connections ensured better gradient flow during training, while the attention mechanism enhanced the model's ability to focus on relevant tumor regions. The segmentation network achieved a Dice coefficient of 0.80 and a Jaccard index of 0.66, outperforming several baseline architectures such as U-Net, VGG16 U-Net, and Swin Transformer-based models.

In the classification stage, the segmentation output was used to enhance the discriminative capacity of the classifier. Each MRI image was overlapped with its corresponding predicted tumor mask to form a composite representation, effectively guiding the model's focus toward the tumor region. The resulting two-channel input was expanded to three channels and processed using a fine-tuned InceptionV3 network. This segmentation-guided classification strategy enabled the network to learn from both intensity-based and spatially localized features.

The InceptionV3-based model attained a classification accuracy of 93.23%, with a precision of 0.92, recall of 0.91, and F1-score of 0.92. The model showed consistent convergence behavior, as reflected by the smooth decline in loss curves and stable validation accuracy throughout training. Comparative analysis with standalone architectures such as ResNet, DenseNet, EfficientNetV2S, and ConvNeXt-Tiny further validated the advantage of the segmentation-guided approach.

The integration of segmentation and classification tasks within a unified framework not only improved accuracy but also enhanced model interpretability, as the segmentation masks effectively served as attention maps highlighting regions of diagnostic relevance. This dual-stage design reduces dependency on manual annotation and provides a more clinically interpretable decision-making process for radiologists

References

[1] Sachin Gupta, Narinder Singh Punn, Sanjay Kumar Sonbhadra, and Sonali Agarwal. MAG-Net: Multi-task attention guided network for brain tumor segmentation and classi cation. https://doi.org/10.48550/arXiv.2107.12321

[2] Müller, S., Weickert, J., & Graf, N. (2016). Automatic brain tumor segmentation with a fast Mumford-Shah algorithm. In M. A. Styner & E. D. Angelini (Eds.), Medical Imaging 2016: Image Processing (Vol. 9784, pp. 97842S). SPIE.

[3] Zhang, J., Jiang, Z., Dong, J., Hou, Y., & Liu, B. (2020). Attention Gate ResU-Net for automatic MRI brain tumor segmentation. IEEE Access, 8, 58533–58545. https://doi.org/10.1109/ACCESS.2020.2983075

[4] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. https://doi.org/10.48550/arXiv.1505.04597

[5] Akib Mohammed Khan, Alif Ashrafee, Fahim Shahriar Khan, Md. Bakhtiar Hasan, Md. Hasanul Kabir. AttResDU-Net: Medical Image Segmentation Using Attention-based Residual Double U-Net . https://doi.org/10.48550/arXiv.2306.14255

[6] Ahamed, Md & Islam, Md & Hossain, Tahmim & Syfullah, Md Khalid & Sarkar, Ovi. (2023). Classification and Segmentation on Multi-regional Brain Tumors Using Volumetric Images of MRI with Customized 3D U-Net Framework. 10.1007/978-981-19-7528-8_18.

[7] Fazli Wahid, Yingliang Ma, Dawar Khan, Muhammad Aamir, Syed U. K. Bukhari. Biomedical Image Segmentation: A Systematic Literature Review of Deep Learning Based Object Detection Methods. https://doi.org/10.48550/arXiv.2408.03393

[8] Matthew, A., Tan, K., Suharyo, P. B., & Gunawan, A. A. S. (2022). A systematic literature review: Image segmentation on brain MRI image to detect brain tumor. 2022 International Conference on Science and Technology (ICOSTECH), 1–5. https://doi.org/10.1109/ICOSTECH54296.2022.9829177

[9] Zeineldin, R. A., Karar, M. E., Coburger, J., Wirtz, C. R., & Burgert, O. (2020) . DeepSeg: Deep neural network framework for automatic brain tumor segmentation using magnetic resonance FLAIR images. International Journal of Computer Assisted Radiology and Surgery, 15(6), 909–920. https://doi.org/10.1007/s11548-020-02186-z

[10] Irmak, E. Multi-Classification of Brain Tumor MRI Images Using Deep Convolutional Neural Network with Fully Optimized Framework. Iran J Sci Technol Trans Electr Eng 45, 1015–1036 (2021). https://doi.org/10.1007/s40998-021-00426-9

[11] Nazir, M., Shakil, S., & Khurshid, K. (2021). Role of deep learning in brain tumor detection and classification (2015 to 2020): A review. Computerized Medical Imaging and Graphics, 91, 101940. https://doi.org/10.1016/j.compmedimag.2021.101940

[12] Lamrani, D., Cherradi, B., El Gannour, O., Bouqentar, M. A., & Bahatti, L. (2022). Brain tumor detection using MRI images and convolutional neural net

[13] Mehmood, Y., & Bajwa, U. I. (2024). Brain tumor grade classification using the ConvNext architecture. Digital Health, 10, 1–13. https://doi.org/10.1177/20552076241284920

[14] Jiang, L., Yuan, B., Ma, W., & Wang, Y. (2023). JujubeNet: A high-precision lightweight jujube surface defect classification network with an attention mechanism. Frontiers in Plant Science, 13, 1108437. https://doi.org/10.3389/fpls.2022.1108437

[15] Nillmani, Sharma, N., Saba, L., Khanna, N. N., Kalra, M. K., Fouda, M. M., & Suri, J. S. (2022). Segmentation-based classification deep learning model embedded with explainable AI for COVID-19 detection in chest X-ray scans. Diagnostics, 12(9), 2132. https://doi.org/10.3390/diagnostics12092132

[16] Mohammad Zafer Khaliki & Muhammet Sinan Başarslan. Brain tumor detection from images and comparison with transfer learning methods and 3-layer CNN. volume 14, Article number: 2664 (2024). Scientific Reports https://doi.org/10.1038/s41598-024-52823-9

[17] Nurul Fatihah Binti Ali, Siti Salasiah Mokri, Syahirah Abd Halim, Noraishikin Zulkarnain, Ashrani Aizuddin Abd Rahni and Seri Mastura Mustaza, "Comparison Review on Brain Tumor Classification and Segmentation using Convolutional Neural Network (CNN) and Capsule Network" International Journal of Advanced Computer Science and Applications(IJACSA), 14(4), 2023. Comparison Review on Brain Tumor Classification and Segmentation using Convolutional Neural Network (CNN) and Capsule Network

[18] McDermott, J. (n.d.). *Hands-on Transfer Learning with Keras and the VGG16 Model*. https://www.learndatasci.com/tutorials/hands-on-transfer-learning-keras/

[19] [19] Nazir, M., Ali, M. J., Tufail, H. Z., Shahid, A. R., Raza, B., Shakil, S., & Khurshid, K. (2022). *Brain Tumor Detection and Segmentation Using Multi-Task Learning Architecture with U-Net.* Journal of Electronic Imaging, 31(5), 051606. https://doi.org/10.1117/1.JEI.31.5.051606

[20] [20] Gupta, S., Kumar, S., & Singh, P. (2021). *MAG-Net: A Multi-Task Attention Guided Network for Brain Tumor Segmentation and Classification.* Computers in Biology and Medicine, 134, 104459. https://doi.org/10.1016/j.compbiomed.2021.104459

[21] [21] Huang, Y., Chen, H., Liu, Z., & Yang, X. (2021). *Deep Multi-Task Learning Framework Based on V-Net for Brain Tumor Segmentation.* Medical Image Analysis, 73, 102169. https://doi.org/10.1016/j.media.2021.102169

[22] [22] Sobhaninia, Z., Ahmadi, M., Hosseinzadeh, M., & Ghafoorian, M. (2023). *Cascaded Multi-Task Network for Tumor Localization, Segmentation and Classification in MRI.* Diagnostics, 13(4), 811. https://doi.org/10.3390/diagnostics13040811

[23] [23] Rabby, A., Rahman, S., & Chowdhury, R. (2024). *Integrated Segmentation-Guided Classification Framework for Brain Tumor Recognition.* Scientific Reports, 14, 102. https://doi.org/10.1038/s41598-024-73055-x

[24] [24] Oktay, O., Schlemper, J., Folgoc, L. L., et al. (2018). *Attention U-Net: Learning Where to Look for the Pancreas.* arXiv preprint arXiv:1804.03999. https://doi.org/10.48550/arXiv.1804.03999

[25] [25] Zhang, Y., Zhang, Q., & Zhao, Y. (2020). *Attention Gate Residual U-Net for Automatic MRI Brain Tumor Segmentation.* IEEE Transactions on Medical Imaging, 39(11), 3583–3596. https://doi.org/10.1109/TMI.2020.3006488

[26] [26] Ngo, M. T., Nguyen, H. Q., & Pham, C. H. (2020). *Multi-Task Learning with Feature Reconstruction for Small Tumor Segmentation.* Pattern Recognition Letters, 138, 325–333. https://doi.org/10.1016/j.patrec.2020.08.002

[27] [27] Li, Z., Wang, J., & Zhou, Y. (2023). *A Multi Brain Tumor Region Segmentation Model Based on Enhanced 3D U-Net.* Applied Sciences, 13(16), 9282. https://doi.org/10.3390/app13169282

[28] [28] Qamar, S., Ahmad, P., & Shen, L. (2020). *HI-Net: Hyperdense Inception 3D U-Net for Brain Tumor Segmentation.* arXiv preprint arXiv:2012.06760. https://doi.org/10.48550/arXiv.2012.06760

[29] [29] Lv, C., Wang, L., & Yu, B. (2025). *BrainTumNet: Multi-Task Deep Learning Framework for Brain Tumor Segmentation and Classification.* Bioengineering, 12(3), 311. https://doi.org/10.3390/bioengineering12030311

[30] [30] Odong, O. J. P., Alemayehu, G., & Nweke, H. F. (2025). *An Integrated Deep Convolutional Neural Network for Tumor Segmentation and Classification.* SN Applied Sciences, 7(2), 196. https://doi.org/10.1007/s42452-025-06968-5

[31] Tomar, N. (2021, September 25). *What is Residual Network or ResNet?* Idiot Developer. https://idiotdeveloper.com/what-is-residual-network-or-resnet/

[32] *GoogLeNet: Revolutionizing Deep Learning with Inception*. (n.d.). viso.ai. https://viso.ai/deep-learning/googlenet-explained-the-inception-model-that-won-imagenet/