

Brain Tumor Detection and Segmentation Using U-Net Architecture

Yashashvini Rachamallu,

Department of Computer Science and Engineering, Michigan State University

rachama2@msu.edu

Abstract: Detecting and segmenting brain tumors is a vital aspect of medical imaging, critical for timely diagnosis and effective treatment planning. Traditional methods for identifying tumors in MRI scans can be slow, error-prone, and reliant on radiologist expertise. This project introduces a two-step approach to improve accuracy and efficiency. First, a Convolutional Neural Network (CNN) successfully classifies MRI images with a 97% accuracy rate in detecting whether a tumor is present. If a tumor is detected, the U-Net architecture, a neural network optimized for biomedical image segmentation, is employed to delineate tumor boundaries with a segmentation accuracy of 93%. The model is trained on the Br35H :: Brain Tumor Detection 2020 dataset, ensuring high reliability. This system not only enhances diagnostic accuracy and speed but also serves as a valuable tool to support radiologists, especially in areas with limited access to specialized medical professionals. The proposed approach demonstrates significant potential in advancing the precision and efficiency of brain tumor diagnostics.

1 Introduction

Detecting and segmenting brain tumors is crucial in modern medical imaging, as early diagnosis and accurate treatment planning can greatly impact patient survival rates. Manual detection from MRI scans is not only labor-intensive but also susceptible to human error and reliant on the expertise of radiologists. Brain tumors can vary significantly in shape, size, and location, making their precise segmentation a difficult task. The U-Net architecture, a convolutional neural network specifically designed for biomedical image segmentation, has proven to be an effective solution for these challenges. Its encoder-decoder structure allows for the capture of both high-level contextual features and detailed information, enabling pixel-level accuracy in defining tumor boundaries.

1.1. Why is it important?

Detecting and segmenting brain tumors is crucial for early diagnosis and effective treatment, directly improving patient survival rates. Brain tumors vary widely in shape, size, and location, making them challenging to diagnose manually. Traditional analysis of MRI scans is time-intensive, prone to errors, and relies heavily on radiologist expertise, leading to potential delays, especially in underserved areas. This project addresses these issues with a two-step automated approach. A Convolutional Neural Network (CNN) first detects whether a tumor is present, filtering scans for further analysis. If detected, the U-Net architecture segments the tumor with pixel-level precision, leveraging its encoder-decoder design for accurate boundary delineation. This system enhances diagnostic reliability, reduces radiologist workload, and ensures timely, accurate results, particularly in regions with limited medical resources.

1.2. Real-World Applications:

- Clinical Diagnosis: Automated tumor segmentation speeds up the diagnostic process, assisting radiologists in planning treatments.
- Telemedicine: Remote diagnostic systems can support healthcare providers in underserved regions.
- Research: Facilitates large-scale studies by efficiently processing extensive MRI.
- Personalized Medicine: Aids in targeted therapies.

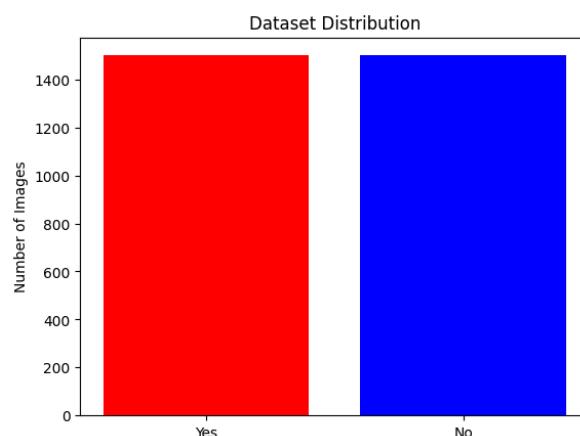
2 Dataset

The dataset for this project, sourced from [Kaggle's Brain Tumor Detection Dataset](#) comprises MRI brain images curated to facilitate tumor detection and segmentation tasks. The data is classified into two main categories: tumorous and non-tumorous. These labeled images followed by annotations of tumour are essential for training deep learning models to effectively distinguish between tumor regions and healthy brain tissue in MRI scans.

2.1 Classification Task

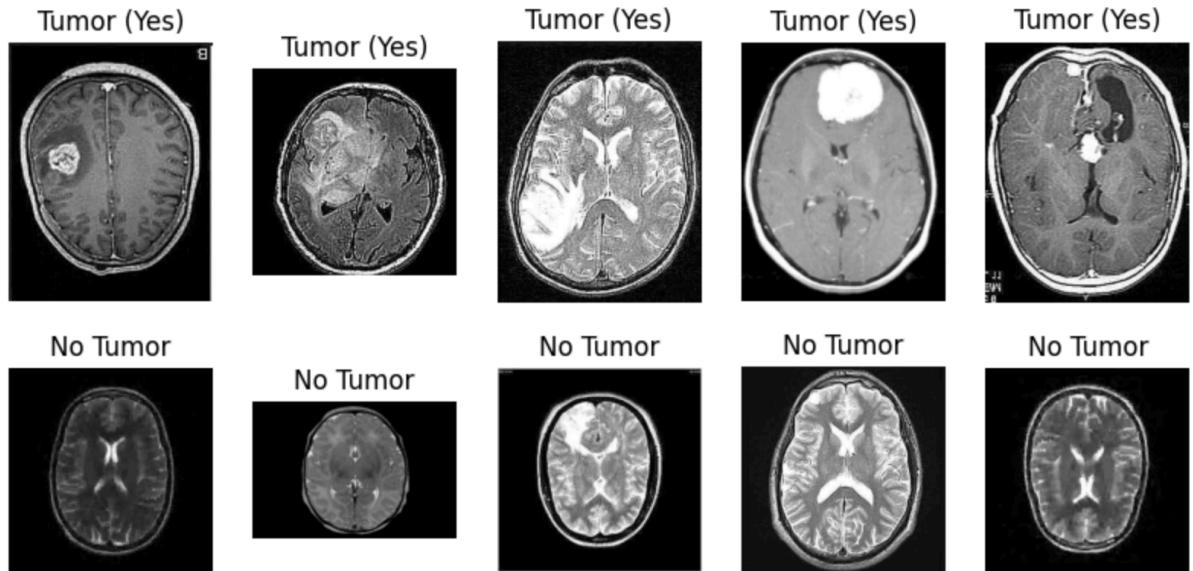
The dataset comprises 3,060 brain MRI images, organized into the following categories:

1. Tumorous Images:
 - a. Number of images: 1,500
 - b. These images display various types of brain tumors, including irregular shapes, sizes, and locations.
2. Non-Tumorous Images:
 - a. Number of images: 1,500
 - b. These images depict healthy brain scans without any detectable abnormalities.
 - c. They serve as critical counterexamples to help the model learn to distinguish between normal and pathological tissue.



The above dataset distribution plot illustrates the number of images categorized into two primary classes: Yes (images containing tumors) and No (images without tumors). The

dataset is well-balanced, with both categories having a comparable number of images, ensuring unbiased training for the classification task. This balance is crucial for building a robust deep learning model capable of accurately identifying the presence or absence of tumors in MRI scans. Below is a sample image from the dataset, which highlights the clarity and quality of the MRI scans used for training. This ensures the model has access to high-resolution and well-annotated images for effective learning.



2.2 Segmentation Task

The dataset for the segmentation task is organized into three categories, each containing images and corresponding annotation files:

1. Training Data:

- **Number of Images:** 500 (increased to 2,000 through augmentation)
- **Description:** The training dataset consists of annotated brain MRI images used to train the segmentation model. The annotations include two shape types: polygons (for irregular shapes) and ellipses (for circular or elliptical regions).
- **Augmentation Techniques:**
 - Horizontal Flip
 - Rotation by 30 Degrees
 - Vertical Flip

These techniques were applied to expand the dataset size, ensuring better model generalization.

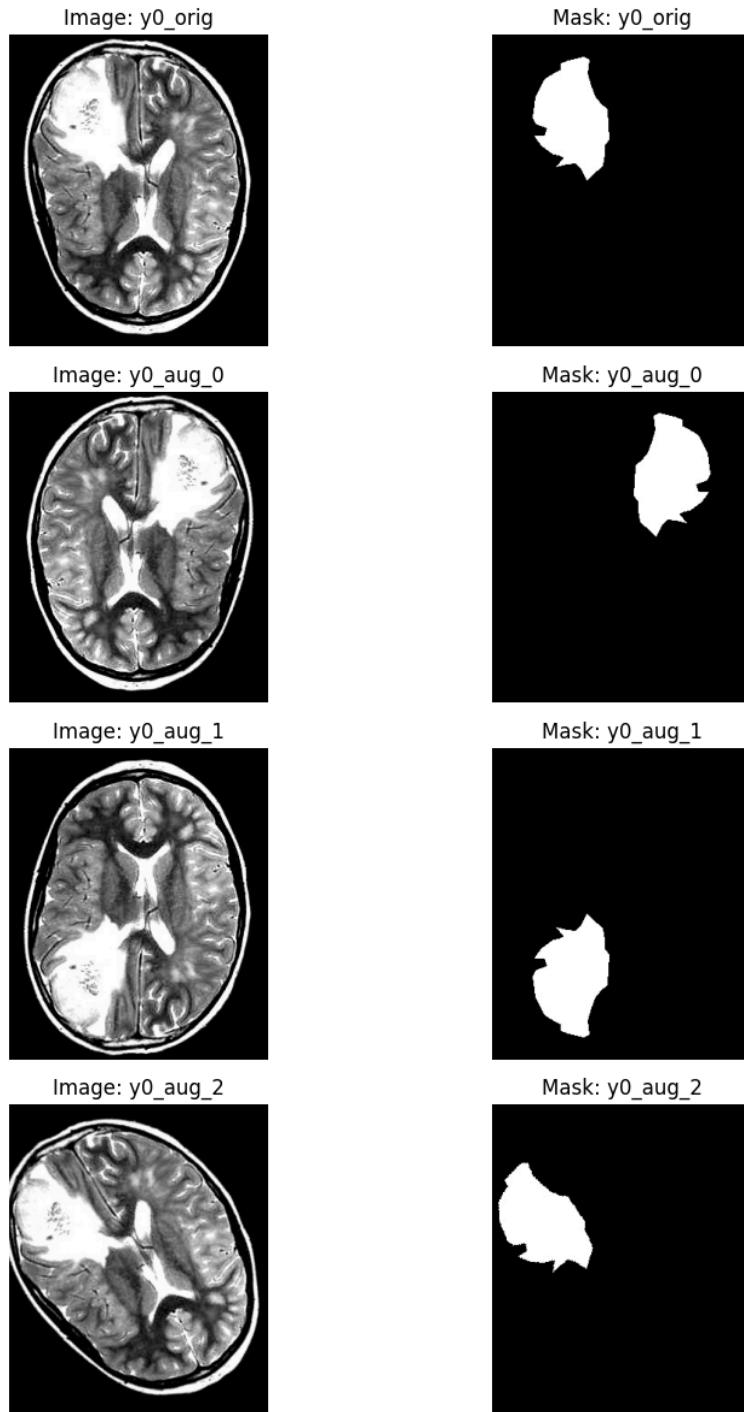
2. Validation Data:

- **Number of Images:** Maintained as per the original dataset
- **Description:** The validation set contains annotated MRI images used to monitor the model's performance during training. No augmentation was applied to preserve the dataset's integrity for validation.

3. Test Data:

- **Number of Images:** Maintained as per the original dataset
- **Description:** The test set includes annotated MRI images for final evaluation of the model's accuracy. Similar to the validation data, no augmentation was performed on the test set to ensure unbiased performance evaluation.

The annotated images in the training, validation, and test datasets play a critical role in enabling the segmentation model to identify tumor regions with high accuracy. Below is a sample image from the augmented training dataset, demonstrating the annotations applied during preprocessing.



2.3 Data Format and Dimensions

- Image format: The images are in grayscale format, a standard for MRI scans, where pixel intensity represents tissue characteristics.
- Resolution: The images have been resized to a uniform resolution of 256x256 pixels to ensure compatibility with the CNN and U-Net architecture and to reduce computational overhead.
- Annotations: Tumorous images are accompanied by binary masks, where pixels labeled as “1” represent tumor regions, and pixels labeled as “0” represent the background and healthy tissue.

2.4 Data Preprocessing

- Data Loading: Images are loaded from the dataset folders using Python libraries like cv2 and organized into labeled arrays for processing. Each image is paired with its respective label (1 for tumorous and 0 for non-tumorous).
- Resizing: To ensure uniformity, all images are resized to 256x256 pixels. This standardization simplifies model input requirements and reduces computational overhead.
- Normalization: Pixel intensity values are scaled to the range [0, 1]. This normalization accelerates model convergence during training and helps the model generalize better by reducing sensitivity to variations in pixel intensities.

3 Description of analysis conducted

3.1 Classification Task

3.1.1 Dataset Preparation

The classification task began with preprocessing the brain MRI images to prepare them for model training. The images were normalized to standardize pixel values, ensuring uniform input data and enhancing the model's ability to learn efficiently. Each image was labeled, with 1 representing the presence of a tumor and 0 indicating no tumor. After preprocessing, the dataset was split into training and validation sets to ensure the model could be evaluated effectively on unseen data during training. This structured dataset preparation was critical in enabling the model to distinguish between tumorous and non-tumorous brain scans accurately.

3.1.2 Model Architecture

The architecture of the Convolutional Neural Network (CNN) was designed to effectively capture and process spatial features in the MRI images. The network consists of multiple layers, each serving a specific purpose in the feature extraction and classification process:

1. Convolutional Layers:

The model includes three convolutional layers, each with a set of filters designed to extract spatial features from the images. The first two layers use 32 filters, while the third layer employs 64 filters. A kernel size of 4 is used for all convolutional operations, with ReLU activation functions applied to introduce non-linearity. These layers help the model detect patterns such as edges, textures, and shapes that are crucial for identifying tumors.

2. Pooling Layers:

MaxPooling layers with a pool size of (2, 2) follow each convolutional layer. These layers downsample the feature maps, reducing their spatial dimensions while retaining essential features. This not only reduces computational complexity but also makes the model invariant to small shifts and distortions in the input images.

3. Dropout Layers:

To prevent overfitting, dropout layers are introduced after each convolutional and fully connected layer. Dropout rates of 0.5, 0.3, and 0.15 are applied, respectively, during training. These layers randomly deactivate a fraction of neurons during each iteration, forcing the network to generalize better and preventing reliance on specific neurons.

4. Flattening Layer:

After the final convolutional and pooling operations, the feature maps are flattened into a one-dimensional vector. This step prepares the data for the fully connected layers, which perform the final classification.

5. Fully Connected Layers:

The architecture includes one fully connected layer with 128 neurons and a ReLU activation function. This layer learns high-level features by combining information from all previous layers. Following this, the output layer contains a single neuron with a sigmoid activation function, producing a binary classification result indicating the presence or absence of a tumor.

Refer to Fig 1 for a detailed visualization of the CNN architecture, illustrating the arrangement of layers and their respective configurations.

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 256, 256, 32)	544
max_pooling2d (MaxPooling2D)	(None, 128, 128, 32)	0
dropout (Dropout)	(None, 128, 128, 32)	0
conv2d_1 (Conv2D)	(None, 128, 128, 32)	16,416
max_pooling2d_1 (MaxPooling2D)	(None, 64, 64, 32)	0
dropout_1 (Dropout)	(None, 64, 64, 32)	0
conv2d_2 (Conv2D)	(None, 64, 64, 64)	32,832
max_pooling2d_2 (MaxPooling2D)	(None, 32, 32, 64)	0
dropout_2 (Dropout)	(None, 32, 32, 64)	0
flatten (Flatten)	(None, 65536)	0
dense (Dense)	(None, 128)	8,388,736
dropout_3 (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 1)	129

Fig 1 Architecture of CNN Model

3.1.3 Training Process

The CNN was trained using a binary cross-entropy loss function and the Adam optimizer, known for its efficiency in optimizing deep learning models. The model was trained for 15 epochs with a batch size of 4, balancing computational requirements and convergence speed. Early stopping was used to halt training once the validation performance plateaued, ensuring the model did not overfit the training data. The validation dataset was crucial for monitoring performance during training, providing an unbiased evaluation of the model's generalization capability. See Fig 2 for an intuition about the training and validation loss trends during the training process.

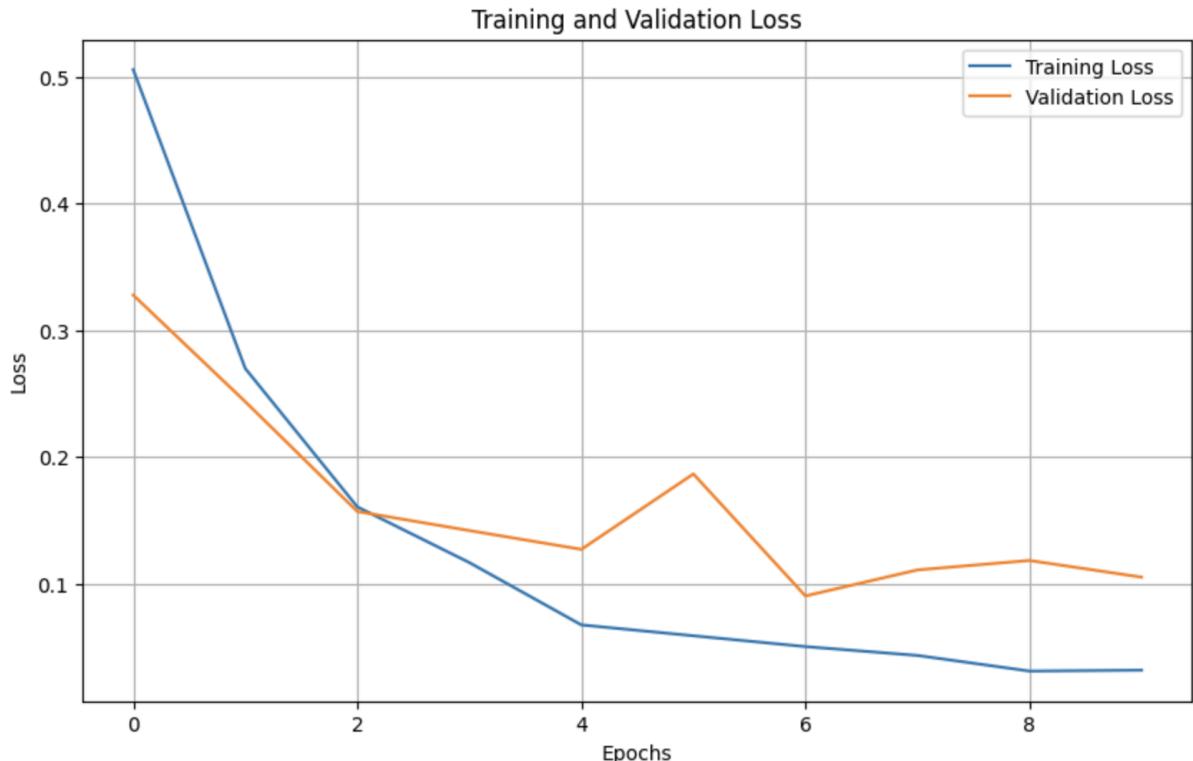


Fig 2

3.2 Segmentation Task

3.2.1 Dataset Preparation

The segmentation task began with preprocessing the brain MRI images to prepare them for training. The images were normalized to standardize pixel values, ensuring uniformity and improving model performance. Data augmentation was then applied to enhance the size and diversity of the dataset, introducing variations such as horizontal flipping, rotations, and vertical flipping. After preprocessing and augmentation, two arrays were created: one containing the original images and the other containing their corresponding annotated images. These annotations represent the tumor regions in the MRI scans, forming the ground truth for the segmentation task. This comprehensive dataset ensured that the model could learn to map raw MRI inputs to their corresponding segmented outputs accurately.

3.2.2 Model Architecture

For the segmentation task, a U-Net architecture was implemented, which is specifically designed for biomedical image segmentation. The U-Net model follows an encoder-decoder structure, enabling it to capture both low-level spatial details and high-level contextual information. The architecture can be summarized as follows:

1. **Encoder Path (Contracting Path):**
 - Consists of multiple convolutional layers with ReLU activation, followed by MaxPooling layers to progressively downsample the spatial dimensions.
 - Dropout layers are incorporated to prevent overfitting and improve generalization.
 - The encoder captures features at different levels, from simple edges to complex patterns.
2. **Bottleneck Layer:**
 - The deepest part of the network contains two convolutional layers with 1,024 filters, followed by a dropout layer.
 - This layer acts as a feature-rich representation of the input image.
3. **Decoder Path (Expanding Path):**
 - Features are progressively upsampled using UpSampling layers, with skip connections concatenating the corresponding feature maps from the encoder.
 - This structure preserves fine-grained spatial information lost during downsampling, enabling precise segmentation.
 - Convolutional layers refine the upsampled features at each stage.
4. **Output Layer:**
 - The final layer consists of a single convolutional filter with a sigmoid activation function, producing a binary segmentation map where each pixel indicates the probability of belonging to a tumor region.

This encoder-decoder structure, combined with skip connections, allows the U-Net to achieve high accuracy in segmenting tumors while maintaining fine boundary details.

3.2.3 Training Process

The U-Net model was trained using the normalized original images as input and their corresponding annotated images as output. The binary cross-entropy loss function was employed to measure the difference between the predicted and true segmentation maps, while the Adam optimizer ensured efficient gradient updates. The training process utilized the following configurations:

- **Learning Rate Scheduler:** A ReduceLROnPlateau callback was used to reduce the learning rate by a factor of 0.5 when the validation loss plateaued, with a minimum learning rate set to 1e-6.
- **Early Stopping:** The training process was halted early if the validation loss did not improve for two consecutive epochs, restoring the best weights for optimal model performance.

The training process was executed over 10 epochs with a batch size of 4, balancing computational efficiency and model convergence. Validation data was used to monitor performance during training and prevent overfitting. The training history shows a consistent decrease in training and validation loss, highlighting the model's ability to learn the

segmentation task effectively. See Fig 3 for an intuition about the training and validation loss trends during the training process.

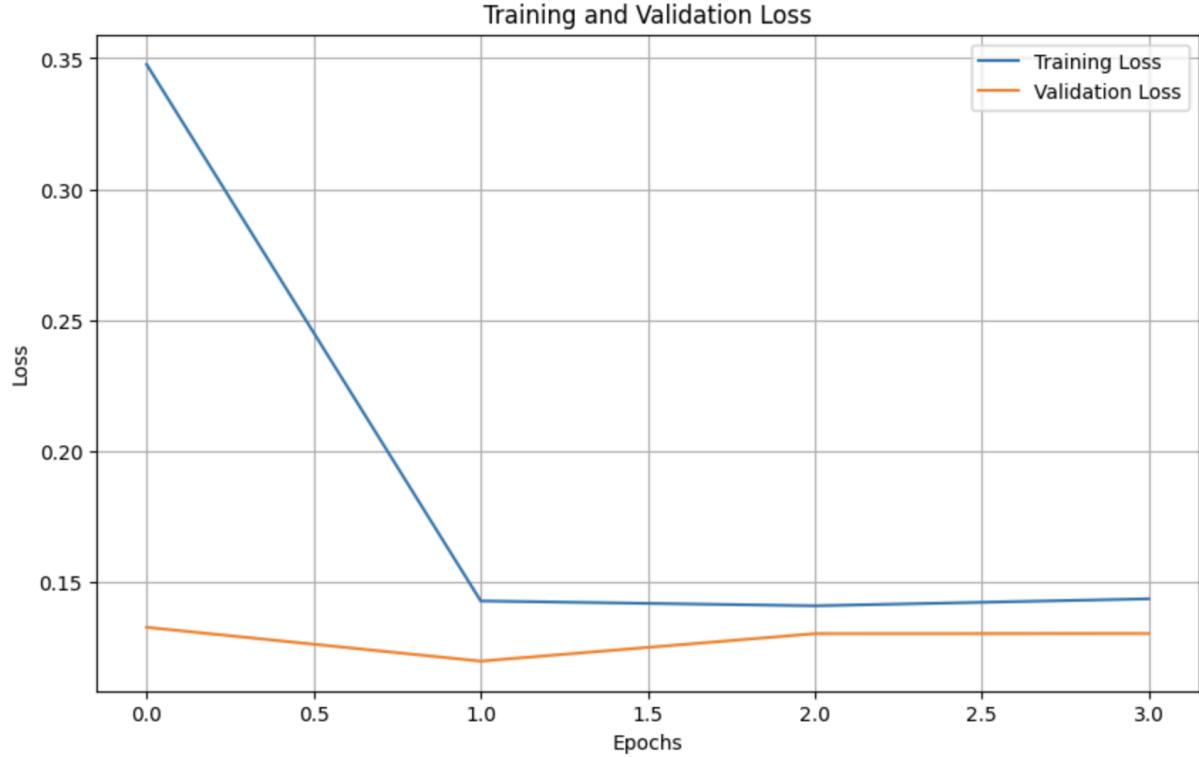
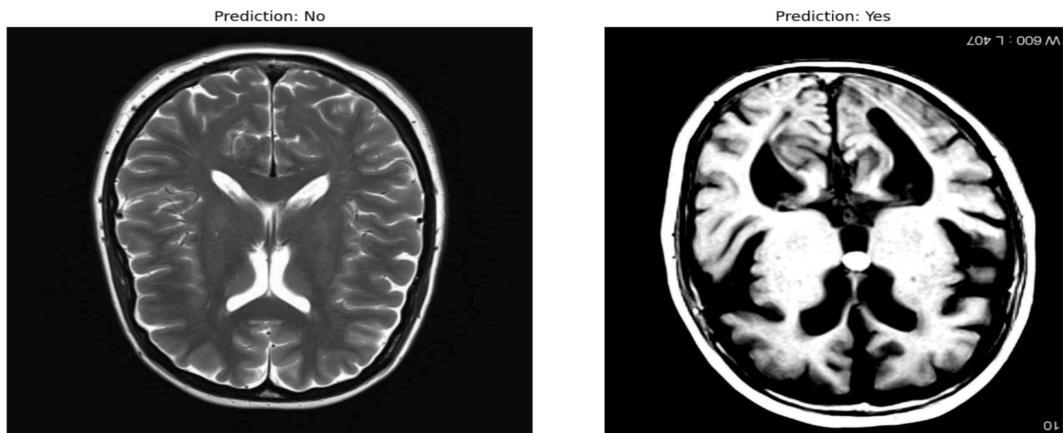


Fig 3

4 Experimentation

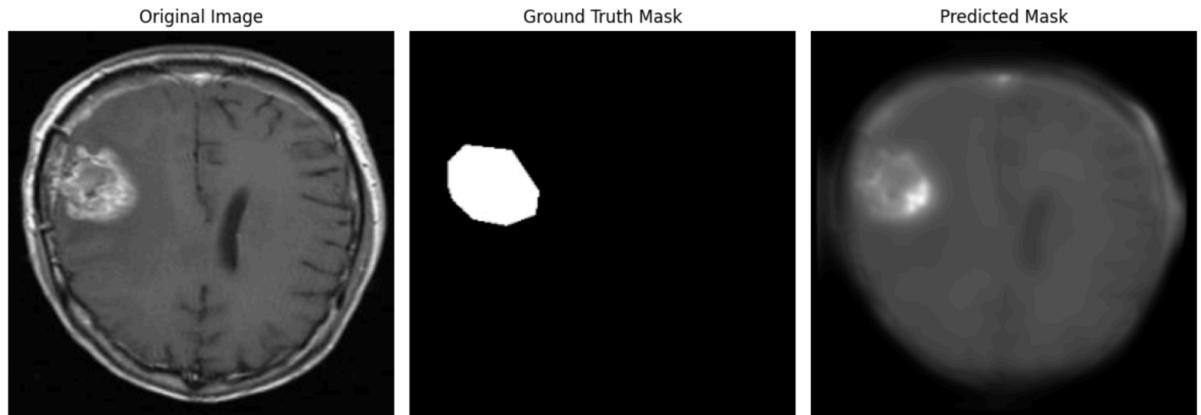
4.1 Classification task

The classification task achieved an accuracy of 97%, demonstrating the model's ability to reliably distinguish between tumorous and non-tumorous brain MRI images. This high level of accuracy is indicative of the effectiveness of the preprocessing steps, dataset preparation, and the carefully designed CNN architecture. The model efficiently learned the patterns and features that differentiate healthy brain tissue from tumor-affected areas, enabling robust and precise predictions. The classification model provides a binary output for each MRI scan, indicating whether the image contains a tumor or not. For example:



4.2 Segmentation task

The segmentation task achieved an accuracy of 93%, showcasing the U-Net model's capability to precisely delineate tumor regions in brain MRI scans. The model effectively utilized the encoder-decoder architecture with skip connections to preserve spatial details, resulting in accurate pixel-level predictions. This accuracy reflects the model's ability to generalize well to unseen data and accurately segment tumor boundaries, even in complex or irregular cases. The segmentation model produces a binary mask for each input MRI scan, where each pixel is assigned a value of 1 (tumor region) or 0 (non-tumorous region).

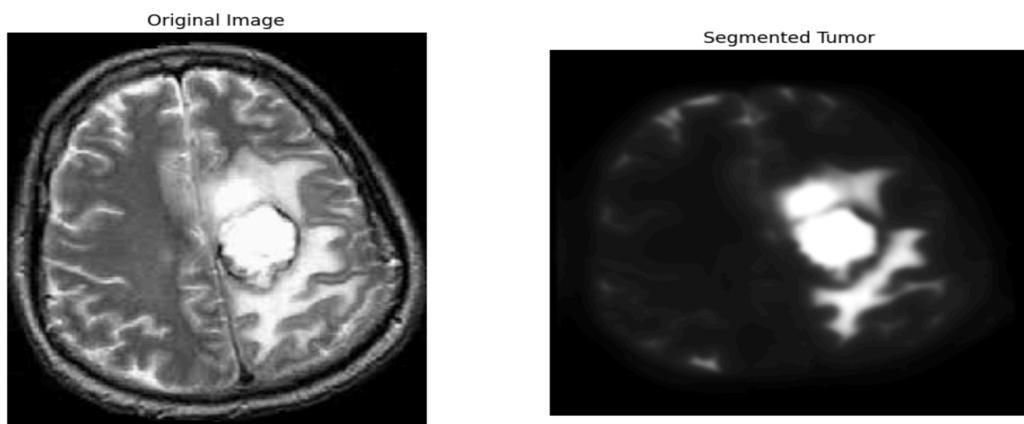


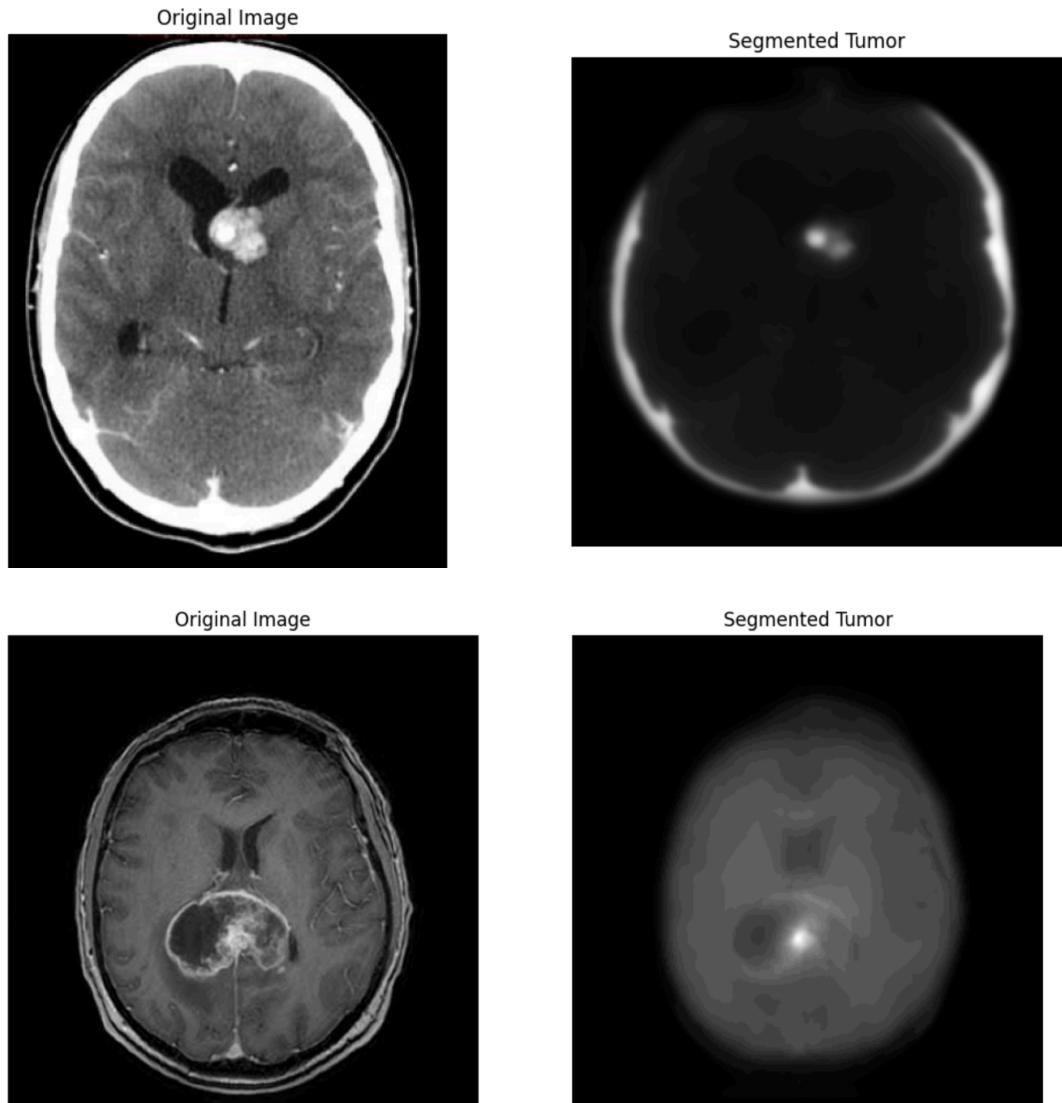
4.3 Testing Procedure

The testing process integrates classification and segmentation to analyze brain MRI scans efficiently. When an MRI scan is input, the system first uses the CNN classification model to determine whether a tumor is present. If the scan is classified as tumorous, the U-Net segmentation model is applied to generate a pixel-wise mask highlighting the tumor's exact location and boundaries. This two-step approach ensures rapid identification of relevant scans and precise segmentation of tumor regions. The final output includes a binary classification result and a segmentation mask, providing radiologists with accurate and actionable insights for diagnosis and treatment planning.

4.4 Results

The following results were obtained using our proposed model for classification and segmentation tasks.





5 Conclusion

This project successfully developed an integrated system for brain tumor detection and segmentation using deep learning models. By combining a CNN for classification and a U-Net for segmentation, the system demonstrates high accuracy in identifying and delineating tumor regions in MRI scans, achieving 97% accuracy for classification and 93% for segmentation. The two-step approach ensures efficient processing by first filtering relevant scans and then providing detailed tumor boundaries, aiding radiologists in accurate diagnosis and treatment planning. While the model shows promising results, challenges such as dataset diversity, segmentation precision for irregular tumors, and real-world adaptability remain.