



ORIGINAL

Brain Tumor Segmentation Pipeline Model Using U-Net Based Foundation Model

Modelo pipeline de segmentación de tumores cerebrales mediante un modelo de base basado en U-Net

Sanjeev Kumar Bhatt¹  , S. Srinivasan¹ , Piyush Prakash²  

¹Research Scholar, Department of Computer Applications, PDM University. Bahadurgarh, Haryana, India.

Cite as: Kumar Bhatt S, Srinivasan DS, Prakash P. Brain Tumor Segmentation Pipeline Model Using U-Net Based Foundation Model. Data and Metadata. 2023;2:197. <https://doi.org/10.56294/dm2023197>

Submitted: 11-8-2023

Revised: 30-10-2023

Accepted: 29-12-2023

Published: 30-12-2023

Editor: Prof. Dr. Javier González Argote 

ABSTRACT

Medical professionals often rely on Magnetic Resonance Imaging (MRI) to obtain non-invasive medical images. One important use of this technology is brain tumor segmentation, where algorithms are used to identify tumors in MRI scans of the brain. The foundation model Pipeline is based on U-Net Architecture to handle medical image segmentation and has been fine-tuned in the research paper to segment brain tumors. The model will be further trained on various medical images to segment images for various bio-medical purposes and used as part of the Generative AI functional model framework. Accurate segmentation of tumors is essential for treatment planning and monitoring, and this approach can potentially improve patient outcomes and quality of life.

Keywords: Segmentation; Forward Process; Reverse Process; Unconditional Image Generation; U-Net; Noise Schedulers; Positional Embedding.

RESUMEN

Los profesionales de la medicina recurren a menudo a la Resonancia Magnética (RM) para obtener imágenes médicas no invasivas. Un uso importante de esta tecnología es la segmentación de tumores cerebrales, en la que se utilizan algoritmos para identificar tumores en exploraciones de IRM del cerebro. El modelo básico Pipeline se basa en la arquitectura U-Net para manejar la segmentación de imágenes médicas y se ha puesto a punto en el trabajo de investigación para segmentar tumores cerebrales. El modelo se seguirá entrenando en varias imágenes médicas para segmentar imágenes con diversos fines biomédicos y se utilizará como parte del marco del modelo funcional de IA Generativa. La segmentación precisa de los tumores es esencial para la planificación y el seguimiento del tratamiento, y este enfoque puede mejorar potencialmente los resultados y la calidad de vida de los pacientes.

Palabras clave: Segmentación; Proceso Hacia Delante; Proceso Hacia Atrás; Generación Incondicional de Imágenes; U-Net; Programadores de Ruido; Incrustación Posicional.

INTRODUCTION

Around the globe, the number of brain tumor cases differs, but it is projected that there are roughly 300 000 new diagnoses every year. The most prevalent kind of primary brain tumor is Gliomas, making up around 80 of all cases. Other frequently occurring types of brain tumors are Meningiomas, pituitary and schwannomas.

Awareness of the various symptoms of brain tumors cause is essential. Depending on the size, location, and type of tumor, symptoms can range from headaches and seizures to changes in personality or behaviour.

Common symptoms include numbness or tingling, vision or hearing problems, and focal weakness. If anyone experiences any of the symptoms, medical attention is necessary to rule out the possibility of a brain tumor from the medical professional.

When diagnosing and evaluating brain tumors, MRI is the most commonly used imaging modality other than imaging tools like (CT) and positron emission tomography (PET). It helps to get detailed images of the brain and surrounding structures, which can accurately localize and characterize the tumor.

Accurate segmentation of brain tumors is crucial for treatment planning and monitoring. However, manual segmentation can be quite time-consuming and labour-intensive, and different observers may need more consistency. Automated segmentation methods can enhance the accuracy and efficiency of tumor segmentation.

The U-Net architecture has demonstrated remarkable efficacy in medical image segmentation through deep learning. It comprises two essential pathways: the encoder pathway to extract input image features and the decoder pathway, which generates the segmented image. U-Net has been widely leveraged to segment brain tumors in MRI images successfully.

Our medical data pipeline incorporates the U-net architecture to efficiently handle segmentation in our model. This pipeline is a crucial component of our broader diffusion model framework, tailored for medical data segmentation. Our paper details a meticulously fine-tuned model to identify and segment brain tumors.

The model will be integrated with the Foundation Generative Pipeline Google Vertex AI cloud platform. This transfer learning approach will unlock the model's potential for real-time segmentation, a critical step towards enhancing its capabilities and making it even more valuable for various applications. The model can be trained with diverse data-sets to allow for customization and adaptation according to the unique needs of medical professionals who use it.

U-Net segmentation boasts several advantages over traditional methods, including its ability to grasp intricate voxel relationships, withstand noise and artifacts, and simplify implementation and training. As such, it is a highly effective tool for image segmentation.

Data-Set

Our team of researchers utilized a cutting-edge model to meticulously examine and scrutinize a comprehensive dataset comprising various brain tumor images, all in the Matlab file format, commonly known as the mat file.

This brain tumor dataset contains 3064 T1-weighted contrast-enhanced images⁽¹⁾ from 233 patients with three kinds of brain tumor: meningioma (708 slices), glioma (1426 slices), and pituitary tumor (930 slices). We split the whole dataset into four subsets and archived them in four compressed files containing 766 slices. The 5-fold cross-validation indices are also provided.

| Task | Brain Tumor Segmentation |
|--------------------|--------------------------------|
| Source | |
| Image Type | T1 contrast-enhanced Brain MRI |
| Image Format | MAT-file(.mat) |
| Dataset Size | 837,77 MB |
| Number of Patients | 233 |
| Number of Images | 3064 |

Detail Of Mri Sequences

Here are some terms of interest which have been used below:

- TE: Time to echo
- TR: Repetition Time
- CSF: Cerebrospinal fluid
- Flair: Fluid-Attenuated Inversion Recover

There are Four Types of MRI sequences:

- Flair: This sequence has very high TE and TR. Abnormalities are highlighted, while normal CSF appears dark.
- T1: This sequence has a short TE and TR. Both tumors and normal CSF appear dark.
- T1CE: This sequence is similar to T1 and also dark but contrast-enhanced.
- T2: This sequence has a high TR and high TE. Both tumors and normal CSF appear bright.

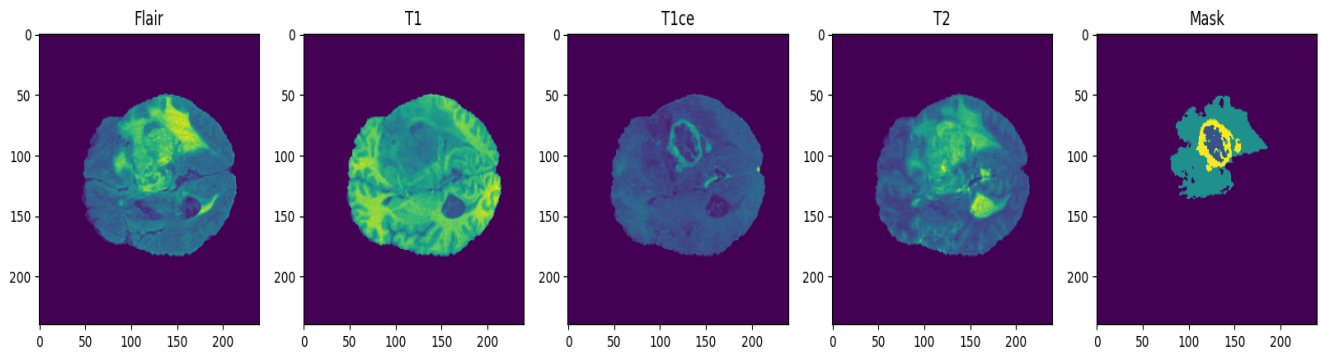


Figure 1. MRI Data Sequence

Segmented: manually selected/segmented Region of interest. This area is where the tumor is situated and serves as the benchmark for validation purposes. It is the ground truth and serves as the benchmark for validation purposes.

DISCUSSION

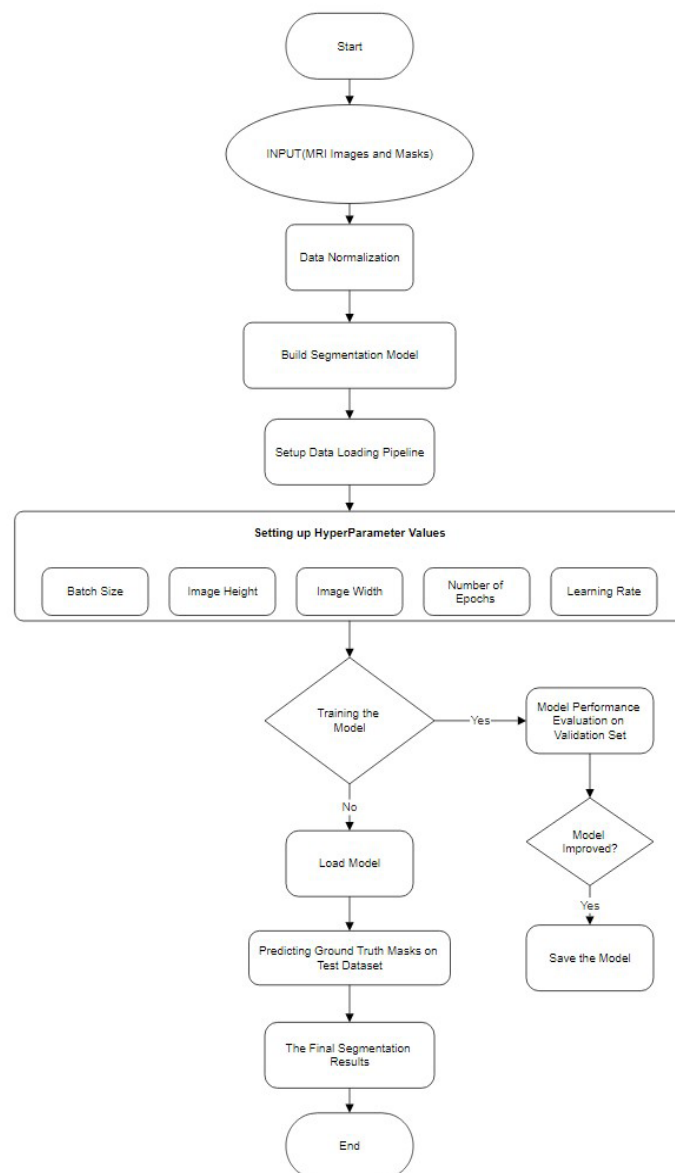


Figure 2. Brain Tumor Pipeline -System Architecture

The U-Net-based architecture employed in this model plays a pivotal role in the framework by efficiently segmenting brain tumors.

Below is a representation of the whole pipeline flow diagram figure 2. Its effectiveness in this task is essential to accurately identify and isolate the affected areas, contributing to a more precise diagnosis and treatment plan. With the aid of Vertex AI, we can seamlessly perform real-time segmentation of bio-medical images. This feature significantly enhances the accuracy and speed of the segmentation process, enabling us to obtain more precise results promptly. Over time, the model will gradually become more malleable and adaptable, enhancing its overall capabilities and performance.

U-Net Architecture

For this study, accurate segmentation is achieved by utilizing a U-net segmentation⁽²⁾ model, initially developed by Olaf Ronneberger for biomedical image segmentation. Our model, which employs the U-net architecture, was developed using PyTorch and has demonstrated effectiveness for this type of task.

The U-Net model can be visualized below figure 3.

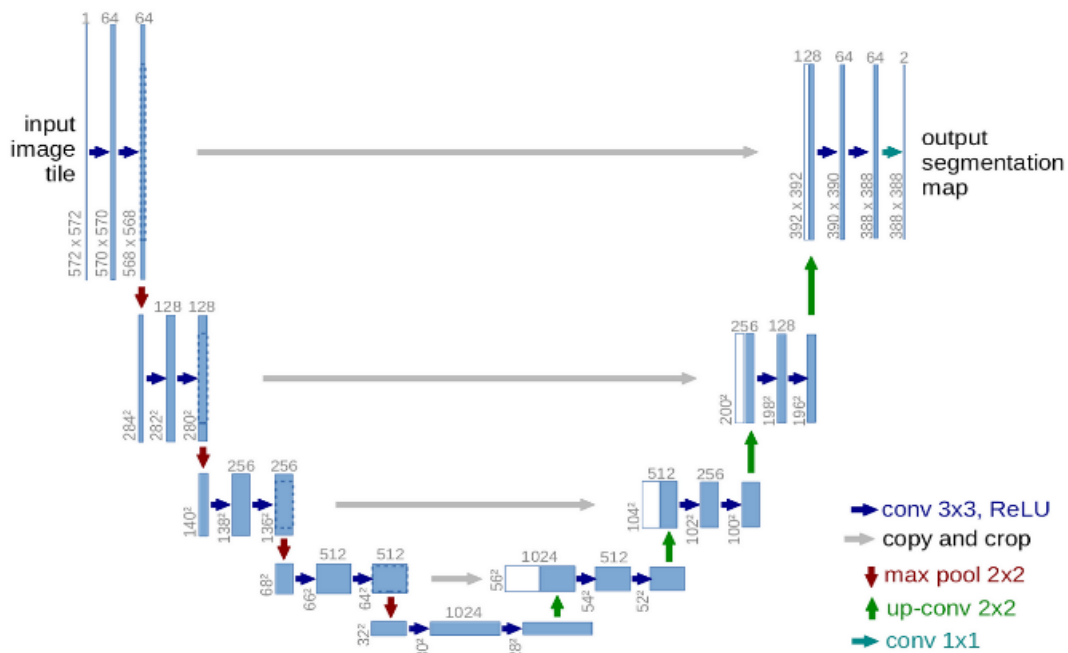


Figure 3. U-Net Architecture

The U-net has an encoder and a decoder. The encoder has two 3 x 3 padded convolutional layers, a batch normalization layer, a ReLU(rectifier linear unit) activation function, and a max-pooling layer of a 2 x 2-sized kernel.

The skip connection connecting two levels in the encoder and decoder part have exact dimensions, as we have applied padding to the convolutions layers. The skip connection connecting two levels in the encoder and decoder part have exact dimensions, as we have applied padding to the convolutions layers.

To ensure that the dimensions and intricacies of images remain intact throughout the various levels of processing, we implement padded layers. By doing so, any potential loss of information during down sampling is accounted for, and the decoder layer is better equipped to restore finer nuances in the image.

U-Net is composed of two paths, a symmetric expanding path (the decoder) and a contraction path (the encoder), totalling 23 convolutional layers.

The encoder or contraction path captures image contexts and is made of repeated stacks of convolution and max-pooling layers. In the encoder, the dimensions of the image gradually decrease.

$$\text{RELU}(q) = \begin{cases} 0, & \text{if } q \leq 0 \\ q, & \text{otherwise} \end{cases} \quad (1)$$

There are 64 filters in the first set of convolutional layers, and in each set after that, the number of filters doubles, for a total of 1024 filters. The deconvolution layer, which is the decoder of the symmetric expanding

path, employs transposed convolutions to localize the ROI in images precisely.

The size of images increases gradually while the depth of the image decreases because of the deconvolution layers. In the decoder, the up-sampling of feature maps commences.

This task is carried out by deconvolution layers of 2×2 , which halve the number of feature channels.

We can produce an initial guess at the ground truth mask by passing in our input images to our U-net model. Initially, our guess will not be excellent, however, we can still use it to compare against our ground truth label.

This comparison gives us an error we can use to adjust our model's parameters, meaning that the next time we pass in an image, we will have a slightly better prediction.

In segmentation, we are trying to learn a mapping from the pixels of an image into the pixels of the segmentation mask. Suppose we have ground truth data such as hand-labelled segmentation masks. In that case, we can train a machine learning model such as the UNET to predict these masks and generalize to new unseen images.

The encoder will extract features from the input image, while the decoder is responsible for up-sampling intermediate features and producing the final output.

Features are passed through an encoder consisting of repeated convolution layers and Max pooling layers that extract intermediate features. These extracted features are then up-sampled by a corresponding decoder, where saved copies of the encoder's features are concatenated onto the decoder features via connecting parts. The final layer produces the output that calculates your loss to a ground truth mask and back-propagates the gradients through the network to improve your model's predictions.

Our model uses a batch normalization at each step. Batch normalization normalizes the inputs to each layer of the network, this keeps the network from over-fitting to the training data. And this leads to faster and more reliable training. It also improves the generalization ability of the network, which means that it is more likely to perform well on unseen data.⁽³⁾

The original paper for U-net did not have this layer as it had not been invented then, but now it is an industry standard and vastly improves the model's performance.

The study uses the Dice coefficient as an accuracy measure for our model, and as the loss function, the optimizer will be Adam, and the learning rate will be 0,001. The dice coefficient is used in a measurement matrix ranging between 0 and 1 to measure the overlap between samples, where a Dice coefficient of 1 denotes perfect and complete overlap.

$$\text{Dice loss} = 1 - 2 * (\text{intersection} / (\text{union} + \text{smooth}))$$

$$\text{DSC} = 2|X \cap Y| / (|X| + |Y|) \quad (2)$$

The loss function used in this study was binary cross entropy. Binary Cross-Entropy loss is a loss function used to measure the difference between two probability distributions. It is typically used in binary classification tasks, where the goal is to predict whether a given input belongs to one of two classes.

The BCE loss function takes a value between 0 and 1, where 0 indicates that the predicted probability perfectly matches the target label, and 1 indicates that the predicted probability is entirely wrong.

$$\text{BCE} = -1/N \sum_{i=0}^N [y_i \cdot \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i)] \quad (3)$$

The accuracy metric we use is Intersection over union (Jaccard Index).

A statistical measure called the Jaccard index, or Jaccard similarity coefficient, assesses how diverse and similar sample sets are. The ratio of verification was created by Grove Karl Gilbert in 1884.

By dividing the measure of the Intersection by the measure of the union of one set, the Jaccard index can be used to assess the degree of similarity between two data sets. Here is how it is defined

$$J(A,B) = |A \cap Y| / (|A \cup Y| + |Y|) \quad (4)$$

Key Attributes of the Foundation Model Pipeline

Traditionally, the U-Net model did not incorporate a batch normalization layer between each level, as Batch Normalization was a relatively new concept when the first U-Net paper was published in 2015. However, our model has successfully implemented the U-Net Architecture and has further improved upon it by incorporating a batch normalization layer. This has led to enhanced performance and accuracy of the model.

The U-Net model is a popular image segmentation technique that utilizes copy and crop functions at each level of the encoder block. This process involves copying the input at each level and passing it to the decoder block to generate better output.^(4,5)

However, this method can result in the loss of some information due to the reduction in image size caused

by the convolutional layers. In order to address this issue, a function is created to copy the image and crop it from the input to the output. This will compensate for the loss of information and help to enhance the accuracy of the output generated by the U-Net model.

When we crop an image, we inevitably lose some information, which can be detrimental to the accuracy of our model. Additionally, the cropping process introduces a degree of randomness that we want to avoid. We can add padding to each convolutional layer to improve the model's performance.

In this case, we will perform zero padding (adding rows and columns of just zero around our real input values) at each convolutional layer so that the image size remains the same after the image goes through the convolutional layers. This will prevent the image size from decreasing at each level, ensuring that the image size remains the same at every level of the encoder and decoder blocks.⁽⁶⁾

By doing so, we can significantly improve our model's overall quality and accuracy. Incorporating padding into the model may result in a minor increase in memory usage, but it ultimately proves advantageous by streamlining the model and retaining essential features.

Models also used early stopping and saving the best model state. Early stopping is an optimization technique that reduces overfitting without compromising model accuracy. The main idea behind it is to stop training before a model starts to over fit after it crosses certain thresholds (patience) when it is training.

The model will also, on each iteration, check its accuracy and save the model state with the ideal amount of loss values automatically.

Why U-net was selected for Building Brain Tumor Foundation Model Pipeline

- Architecture: compared to traditional models, U-net models have a more intricate architecture that enables them to better understand the spatial connections between pixels in an image. This is particularly crucial in segmenting brain tumors, as they may be small and irregular.
- Training data Choosing a suitable machine-learning model is essential when working with medical imaging data. U-Net-based models are a great choice since they are designed to handle the complex structures and high resolutions common in medical images. Traditional deep learning models may only be as effective with fine-tuning to work with medical imaging data.
- Performance: U-Net models have been shown to outperform traditional models on brain tumor segmentation tasks, particularly on datasets with limited training data.
- Small objects segmentation: U-Net models can handle small structural objects in the image, a critical aspect of brain tumor segmentation, where small tumors need to be segmented. Traditional models may not be able to segment small objects as well, as they may not be able to learn the specific features.
- Easy to implement: they are relatively easy to train and deploy. The main parameters that need to be tuned are the learning rate, the number of epochs, and the batch size. The parameters control how quickly the model will be trained, the number of epochs controls how often the model sees the entire training dataset, and the batch size controls how many images are processed.
- Open Source: they are open-source so anyone can use them. The code is well-documented, so it is easy to reproduce the results of the original study^{{cite{b7}}}. This is important for ensuring that the results are reliable.
- Various use cases: They are effective for various other medical image segmentation, like the segmentation of human body organs and tissues.
- Skip connections: U-Net uses skip connections to concatenate features from the encoder and decoder. This helps to preserve spatial information, which is essential for accurate segmentation.

METHODS

Initially, the diffusion model will produce a set of images with the necessary features for effectively segmenting the tumor region. Following this, the U-net model will segment the images and emphasise the area where the tumor is likely situated.⁽⁷⁾

The framework has undergone meticulous fine-tuning to ensure optimal performance in brain tumor segmentation. However, through further calibration and refinement, it is to be adapted for the segmentation of breast cancer images.⁽⁸⁾

As the framework progresses through subsequent iterations, it can dynamically adjust to the data it is presented with and can be efficiently trained to segment a variety of biomedical images accurately.

The U-net model was trained on various epochs, and the learning rate was kept at 0,001 at all the epochs as it is a balanced rate for the U-net architecture.

The algorithm followed can be shown below.

Table 2. Algorithm

Algorithm

1. For each of the training and test images:
 - a. Re-scale all the images such that their dimension is 256X256.
 - b. The MRI Images should be in RGB if needed and the mask Images must be in grayscale.
2. Apply padding to the images, so as to prevent loss of information and keep the size of the image same at each level of encoder and decoder block.
3. Train the model on the 2451 training images over various ranges of epochs and note down their performance.
4. For each of the 2451 training images:
 - a. Predict the mask or the segmentation area.
 - b. Use Binary cross-entropy and Adam Optimizer to calculate the loss values and then optimize the model to maximize the Jaccard Index.
5. Evaluate the model based on the Jaccard Index using a threshold of 0,5
6. For each test data set image, use the optimized model to get the segmented area.
7. Compare the results to the Ground Truth.

RESULTS

Training the network began in small iterations where we monitored the segmentation performance based on iterations alone. Early results using 10 epochs produced poorly.

Poor segmentation results were expected based on the small epoch range used during training. However, adequate results were recorded. This is likely due to the size of each dataset.^(9,10,11)

To further improve these results, we increased the network's epoch range to 50 and monitored the results to see if the segmentation performance had improved. Using 50 epochs significantly improved the network's segmentation performance across all three perspective planes. Training the model on the Multi-modal Brain Tumor Segmentation Challenge 2020⁽³⁾ dataset, we get the following results.

Training Results on U- Net Architecture

The results that we got were as follows after training on T1,T2 and flair , respectively Images with a learning rate of 0,001 with the UNET architecture as follows.

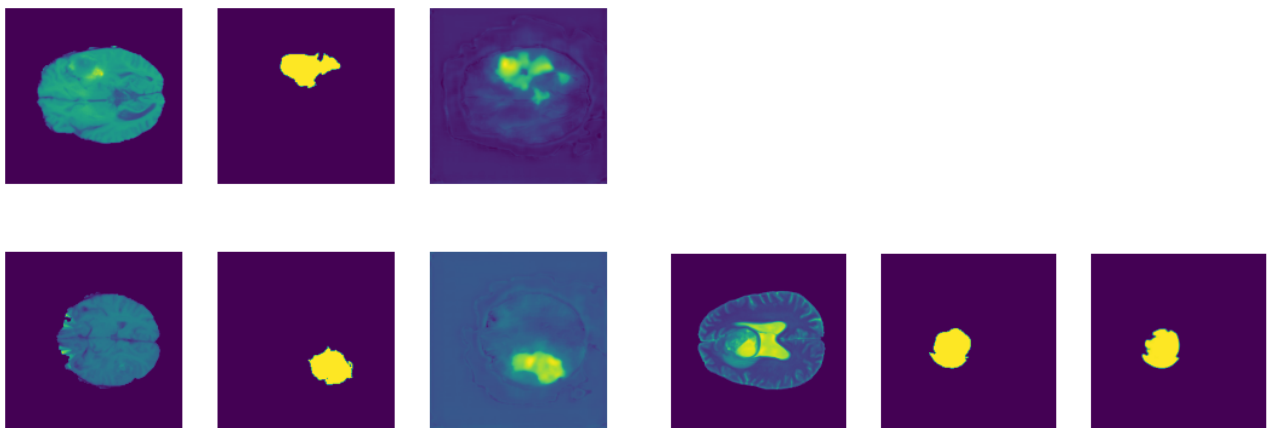


Figure 4. T1

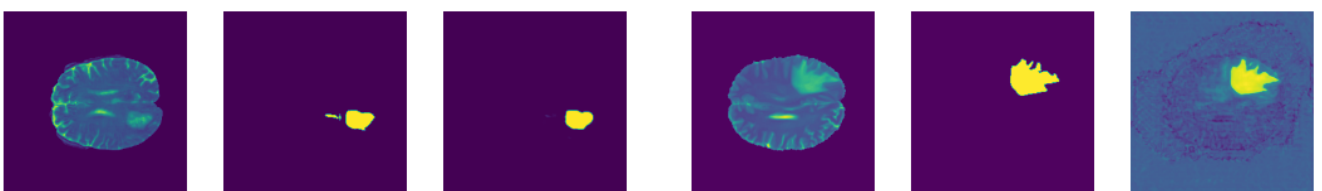


Figure 5. T2

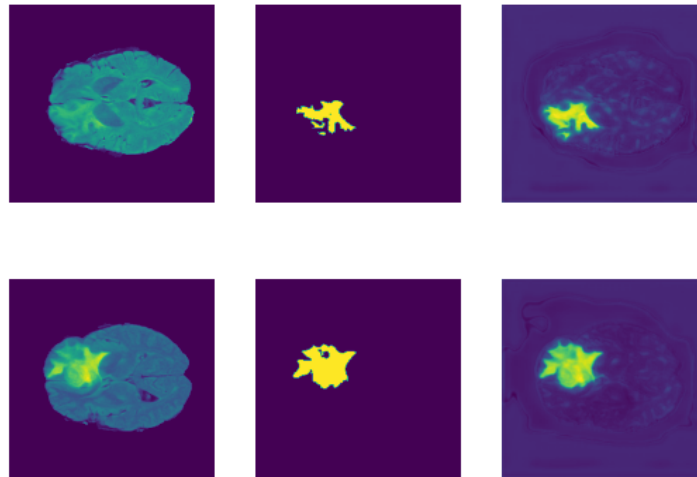


Figure 6. Flair

Foundation Model (Memory Size) - U-Net Based Architecture

Following is the overall memory for model for tumor segmentation.

Table 3. Memory matrix - model

| Dataset | Memory | | | |
|---------|------------------------|-----------------------|------------------------|--------------|
| | Overall Disk Usage(GB) | Overall RAM Usage(GB) | Model memory usage(MB) | Model memory |
| T1 | 4,2 | 5,7 | 828,3 | 516,3676 |
| T2 | 4,3 | 6,0 | 828,3 | 516,3676 |
| T1ce | 4,2 | 5,8 | 828,3 | 513,38 |
| FLAIR | 4,3 | 5,9 | 828,3 | 514,38 |

Table 4. Performance matrix after 200 epocs

| Dataset | Average IOU for 101 EPOCHS | | |
|---------|----------------------------|-----------------------|------------------------|
| | Overall Disk Usage(GB) | Overall RAM Usage(GB) | Model memory usage(MB) |
| T1 | 0,8203 | 0,5728 | 0,5728 |
| T2 | 0,8307 | 0,6563 | 0,6674 |
| T1ce | 0,7799 | 0,5766 | 0,5614 |
| FLAIR | 0,8312 | 0,7630 | 0,7539 |

Foundation Models Results - Brain Tumor segmentation

Step: 1 Results for brain tumor segmentation on Data Set after 26 epocs

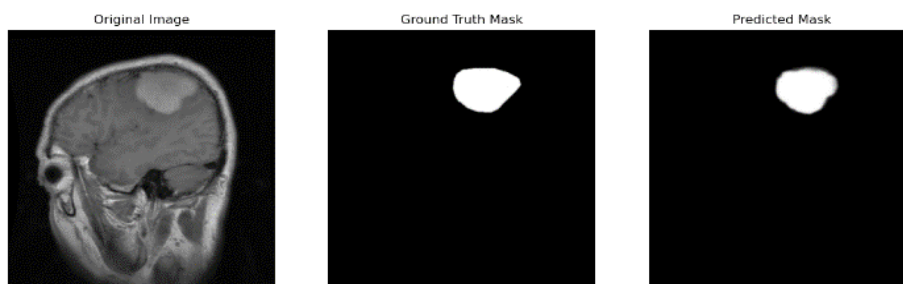


Figure 7. Training Data Set

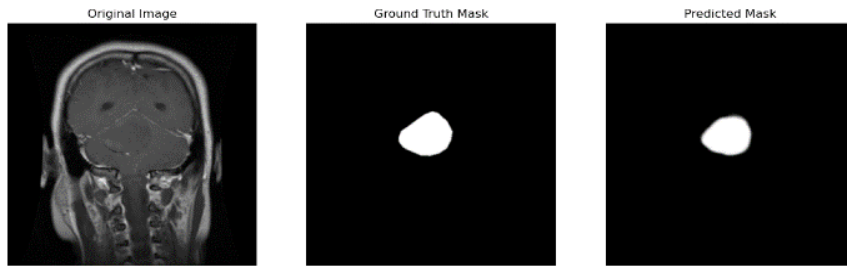


Figure 8. Validation Data Set

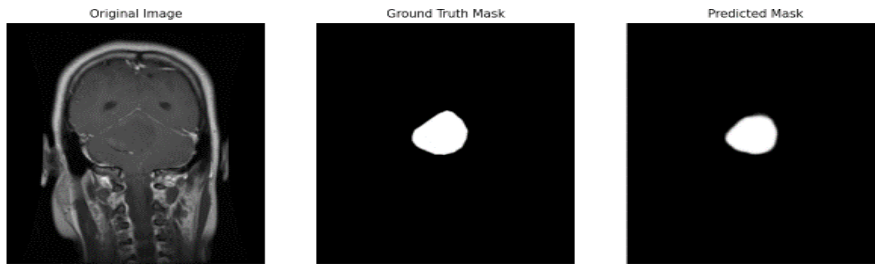


Figure 9. Test set

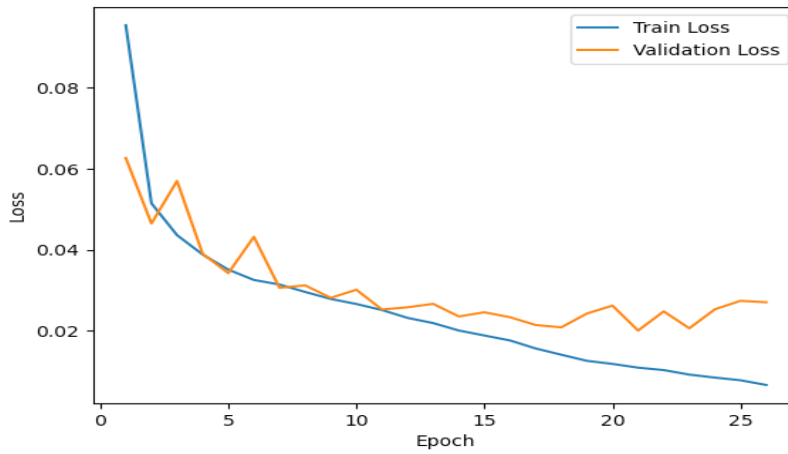


Figure 10. Train and Validation Loss

Step 2: The model was then run at 50 epochs and the results were evaluated. The results are recorded below

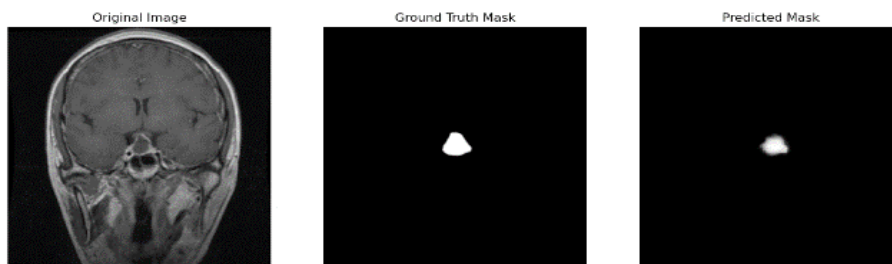


Figure 11. Training Data Set

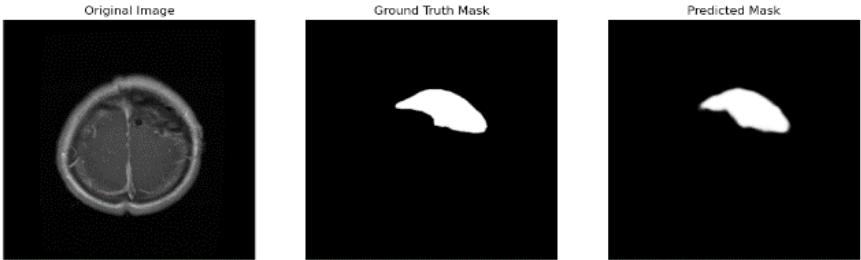


Figure 12. Validation Set

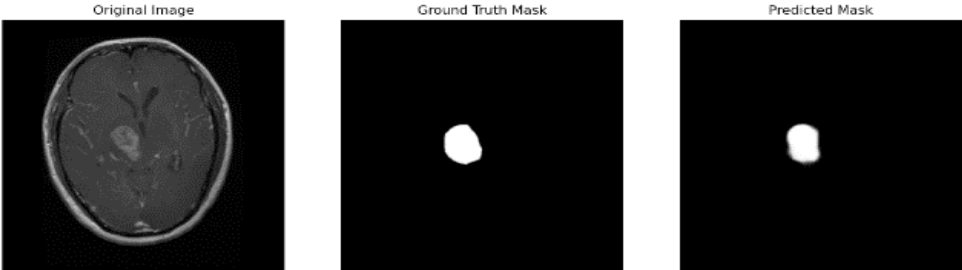


Figure 13. Test set

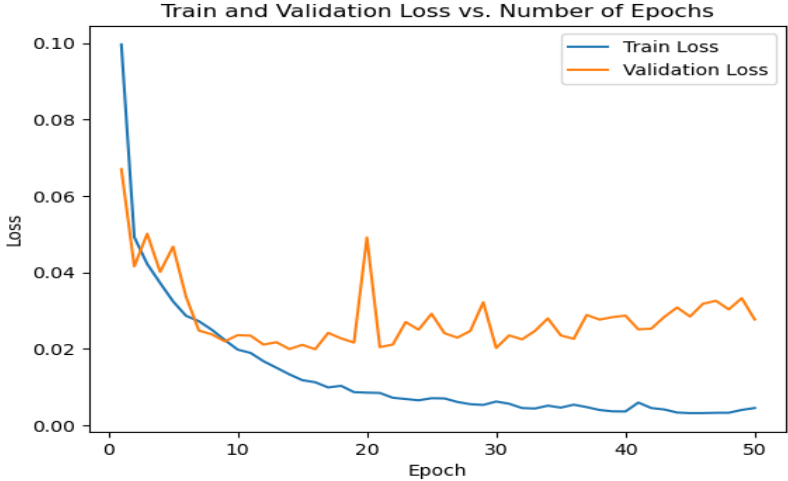


Figure 14. Train and Validation Loss

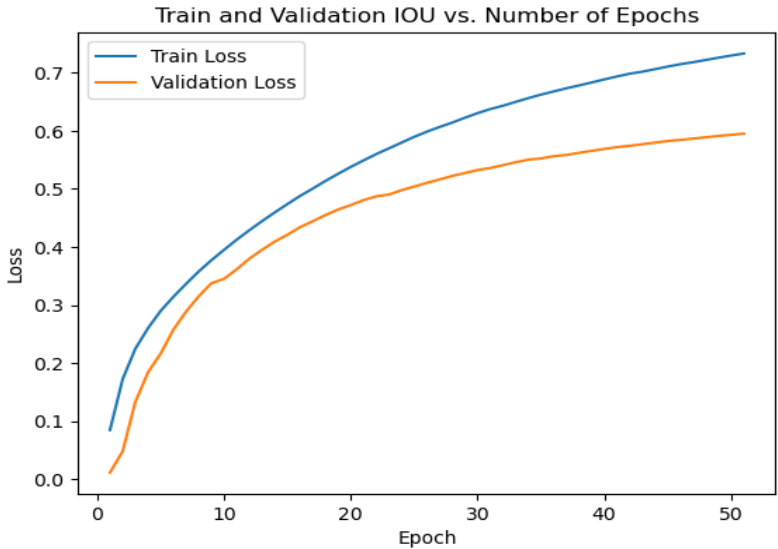


Figure 15. Comparison of epochs vs train and validation IOU

The model we have developed has proven to be highly effective in accurately segmenting tumors from the provided brain tumor dataset and the data generated by the diffusion model. If a tumor feature is absent from an image, the model will display a blank image.

| Epochs | Memory | | | |
|--------|-------------------------|-----------------------|------------------------|----------------|
| | Overall Disk Us-age(GB) | Overall RAM Usage(GB) | Model memory usage(MB) | Model Size(MB) |
| 50 | 4,2 | 2,9 | 494,521 | 118 |
| 151 | 4,3 | 3,5 | 985,897 | 174 |

CONCLUSION

Our model has been trained on reliable datasets sourced from trustworthy sources, and the images generated by the diffusion model contain the same features present in real biomedical images. While our U-Net implementation is currently in its raw form, it will continue to improve over time. Our lightweight network can perform accurate segmentation without aggressive data augmentation. This makes it a valuable tool that trained physicians could use in a medical setting as a secondary evaluator of a patient's MR image.⁽⁴⁾

The application of deep learning in the field of brain tumor segmentation has advanced rapidly, and this study demonstrates how computer vision and deep learning is used in the medical domain and part of the generative AI pipeline to solve to segment brain tumors from two-dimensional MR brain images using a lightweight variant of a well-established architecture.

While further research is necessary to improve the model's performance, this study showcases the potential of deep learning in the medical field to accurately identify and segment bio-medical images.

The framework, when fully integrated, will be able to segment images from various medical datasets and will also be able to perform real-time segmentation using the capabilities of Vertex AI.

REFERENCES

1. P. K. S. B. P. & J. P. C. V. Supe, "Image Segmentation and Classification for Medical Image Processing," <https://core.ac.uk/download/539895621.pdf> , (2019).
2. A. J. S. B. J. K.-. B. H. Menze, ""The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)"" , IEEE Transactions on Medical Imaging 34(10), 1993-2024 (2015).
3. H. A. A. S. M. B. M. R. J. K. e. a. S. Bakas, "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features," Nature Scientific Data, 4:170117 (2017).
4. M. R. A. J. S. B. M. R. A. C. e. a. S. Bakas, "Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge".
5. H. A. A. S. M. B. M. R. J. K. e. a. S. Bakas, "Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection," The Cancer Imaging Archive, 2017.
6. A. Patel, "Benign vs malignant tumors,," JAMA Oncol, 6 (9) (2020) 1488.
7. C. D. M. Ş. A. Işın, "Review of MRI-based brain tumor image segmentation using deep learning methods," Sci. 102 , p. 317-324, (2016).
8. C. D. M. Ş. A. Işın, "Review of MRI-based brain tumor image segmentation using deep learning methods," Procedia Comput ,Sci., p. 317-324, (2016) .
9. M. K. R. A. Md. Sattar, "Automatic cancer detection using probabilistic convergence theory, in: Computational Intelligence in Oncology: Applications in Diagnosis, Prognosis and Therapeutics of Cancers," Springer, p. pp. 111-122, 2022.
10. A. N. M. P. S. D. M.S. Pathan, " Analyzing the impact of feature selection on the accuracy of heart disease prediction,," Healthc. Anal., p. 100060, 2 (2022) .

11. S. K. .. & S. ,. S. .. (. Bhatt, "Lung Cancer Detection Using AI and Different Techniques of Machine Learning.," International Journal of Intelligent Systems and Applications in Engineering, 12(8s), p. 630-638, 2023.

FINANCING

The authors did not receive financing for the development of this research.

CONFLICT OF INTEREST

Declare potential conflicts of interest; otherwise declare "None" or "The authors declare that there is no conflict of interest".

AUTHORSHIP CONTRIBUTION

Conceptualization: Sanjeev Kumar Bhatt, S. Srinivasan, Piyush Prakash.

Data curation: Sanjeev Kumar Bhatt, S. Srinivasan, Piyush Prakash.

Formal analysis: Sanjeev Kumar Bhatt, S. Srinivasan, Piyush Prakash.

Acquisition of funds: Sanjeev Kumar Bhatt, S. Srinivasan, Piyush Prakash.

Research: Sanjeev Kumar Bhatt, S. Srinivasan, Piyush Prakash.

Methodology: Sanjeev Kumar Bhatt, S. Srinivasan, Piyush Prakash.

Project management: Sanjeev Kumar Bhatt, S. Srinivasan, Piyush Prakash.

Resources: Sanjeev Kumar Bhatt, S. Srinivasan, Piyush Prakash.

Software: Sanjeev Kumar Bhatt, S. Srinivasan, Piyush Prakash.

Supervision: Sanjeev Kumar Bhatt, S. Srinivasan, Piyush Prakash.

Validation: Sanjeev Kumar Bhatt, S. Srinivasan, Piyush Prakash.

Display: Sanjeev Kumar Bhatt, S. Srinivasan, Piyush Prakash.

Drafting - original draft: Sanjeev Kumar Bhatt, S. Srinivasan, Piyush Prakash.

Writing - proofreading and editing: Sanjeev Kumar Bhatt, S. Srinivasan, Piyush Prakash.