**AMRITA VISHWA VIDYAPEETHAM**



**DEEP LEARNING**

**22AIE304**

**Research Paper Report**

**TITLE**

**Accurate Brain Tumor Segmentation in Multi-Modal MRI Scans with Optimized Res-U-Net**

**SUBMITTED BY**

**Pula V Lakshmi Narasimha Naidu**  **AV.EN.U4AIE22029**

**Luhya Srinivas AV.EN.U4AIE22013**

**P Bhuvaneshwari AV.EN.U4AIE22030**

**B V B S Pavan Kumar AV.EN.U4AIE22056**

**School of Computing,**

**Amrita Vishwa Vidyapeetham, Amaravati**

**Course Instructor: Dr. M. Rajasekhar Reddy**

**Abstract:**

Brain tumors pose a serious health risk, often leading to significant neurological impairment and contributing to high mortality rates worldwide. Early detection and accurate diagnosis are crucial for improving patient survival rates, with MRI scans playing a central role in identifying brain tumors. A brain tumor forms due to abnormal cell growth—either cancerous or non-cancerous—in a specific region of the brain, and prompt treatment is essential due to the high risks associated with these tumors. Brain tumor segmentation is a critical step in treatment planning and delivering timely medical care.

In this study, we developed and optimized a Res-U-Net-based deep learning model for segmenting brain tumors. The model was trained on the Brain Tumor Segmentation 2020 (BraTS 2020) dataset, which contains labelled MRI scans in various formats. This dataset includes 1,845 MRI scans distributed across 369 training folders, and 500 scans in 125 validation folders. The data was split, with 80% used for training and 20% for testing, and 100% of the validation data reserved for the final evaluation. Each MRI scan is in NIfTI (.nii.gz) format and includes T1, T1ce (contrast-enhanced T1), T2, FLAIR (Fluid Attenuated Inversion Recovery) images, along with manual segmentation labels.

The tumor regions are divided into three main segments: the tumor core, the whole tumor, and the enhancing tumor. The model was trained for 50 epochs, employing nearest neighbor interpolation to manage differences in image resolution. The segmentation accuracy achieved was 99.5% on the training set, 98% on the validation set, and 98.23% on the test set, underscoring the model’s high accuracy and reliability in identifying complex brain tumor structures. These results demonstrate the potential of the Res-U-Net model for supporting effective diagnosis and treatment planning in clinical settings.

**Introduction:**

The human brain is the central part of the nervous system and works alongside the spinal cord to form the Central Nervous System (CNS). It regulates most bodily functions, including the processing, organization, and transmission of instructions to other parts of the body. The structure of the human brain is exceptionally complex, and any disruption within this intricate system, such as the growth of abnormal cells, can lead to conditions like brain tumors.

A brain tumor arises when abnormal cells grow within the rigid confines of the skull, which houses the brain. Glial cells, which provide support to neurons by positioning them and assisting in their functions, are the most prevalent cells in the CNS. Globally, approximately 7 out of every 100,000 individuals are diagnosed with gliomas each year, with glioblastomas being the most common and aggressive form. Glioblastomas predominantly affect men over the age of 45 and account for a large proportion of brain cancer fatalities, with a fatality rate ranging between 18,280 and 18,300 annually due to primary malignant tumors.

To help in the diagnosis and treatment of such tumors, the “Brain Tumor Segmentation (BraTS) 2020” dataset plays a pivotal role in medical imaging research, supporting AI-driven methodologies for brain tumor analysis. This dataset comprises MRI scans of patients with brain tumors, enabling machine learning models to identify and accurately delineate tumor regions. The development of segmentation techniques is -

* **Whole Tumor**: The entire tumor mass, encompassing all abnormal regions.
* **Tumor Core**: The central part of the tumor, often surrounded by edema (swelling).
* **Enhancing Tumor**: Rapidly growing tumor cells that appear enhanced on MRI scans, indicating high metabolic activity.

The BraTS 2020 dataset includes various MRI modalities to capture different aspects of the brain's structure and tumor characteristics. T1-weighted images provide insights into structural contrasts, T2-weighted and FLAIR images highlight fluid accumulation and edema, while T1 contrast-enhanced images reveal areas with increased blood supply, making active tumor regions more discernible. Expert-labeled segmentation masks within the dataset serve as ground truth references, ensuring precise tumor boundary identification.

In this research, we propose two Res-U-Net models designed to identify tumor regions across different MRI scan types. These models aim to enhance the accuracy and efficiency of tumor segmentation.

**Literature Survey:**

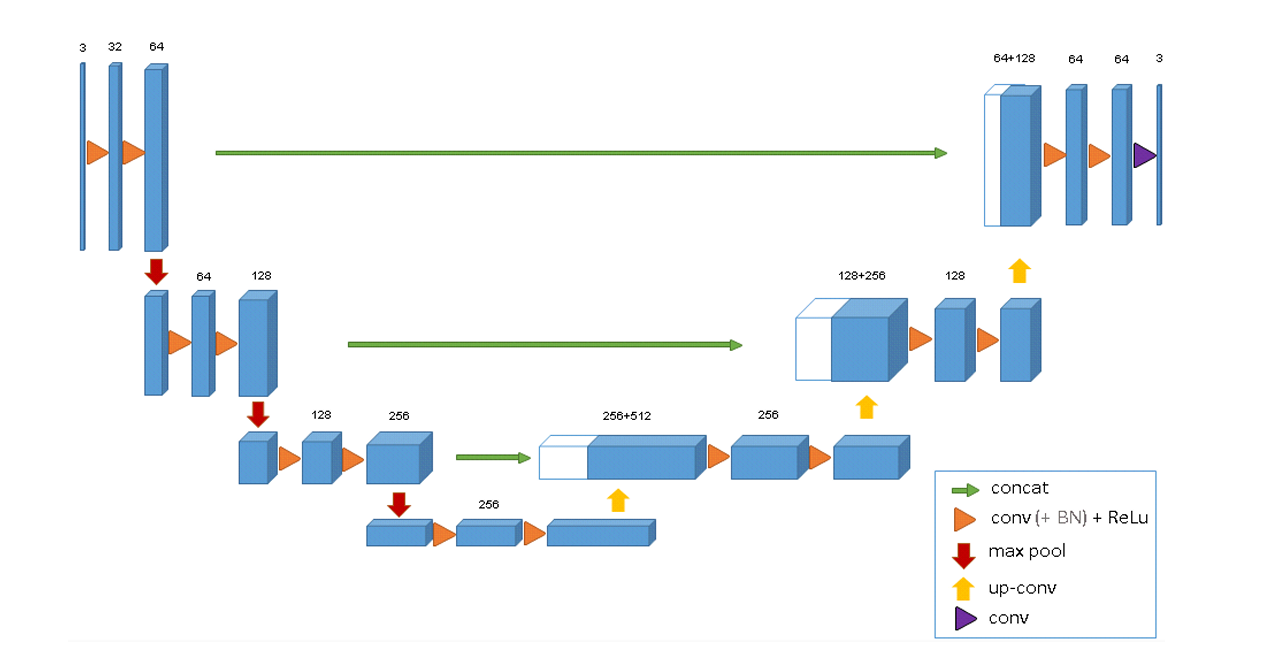
|  |  |  |  |
| --- | --- | --- | --- |
| **Paper** | **Research Gap** | **Efficient model employed** | **Results** |
| Ma et al. [1] | The direct encoder-decoder connection in existing 2D residual models may limit performance, highlighting the need for improved inter-module structures. | deep supervision-based 2D  residual model (a deep supervision-based 2D residual UNet) | Dice Coefficient results of WT, ET, TC are 78.79%, 87.26%, 75.93%. |
| Soltaninejad et al. [2] | The absence of attention or structured nodes in dual encoder-decoder models may limit their performance, particularly in TC detection. This highlights the need for improved methods to bridge semantic gaps effectively. | MR Encoder-Decoder | Dice Coefficient results of WT, ET, TC are 78%, 66%, 70%.  Sensitivity results of WT, ET, TC are 78%, 70%, 74%.  Specificity is 100%. |
| Savadikar et al. [3] | Combining node and attention structures, as implemented in ResUNet+, could potentially outperform models that rely solely on attention blocks for segmentation. | Probabilistic UNet | Dice Coefficient results of WT, ET, TC are 79.88%, 72.49%, 77.71%.  Sensitivity results of WT, ET, TC are 87.9%, 75.13%, 78.13%.  Specificity results of WT, ET, TC are 97.79%, 99.96%, 99.96%. |
| Zhang et al. [9] | The 2D model proposed in BraTS 2019 may lose contextual information across slices, underscoring the need for methods that preserve 3D context in segmentation tasks. | Attention Guided Residual U-Net | Dice Coefficient results of WT, ET, TC are 77.7%, 87%, 70.9%. |

**Existing Techniques used by some Authors:**

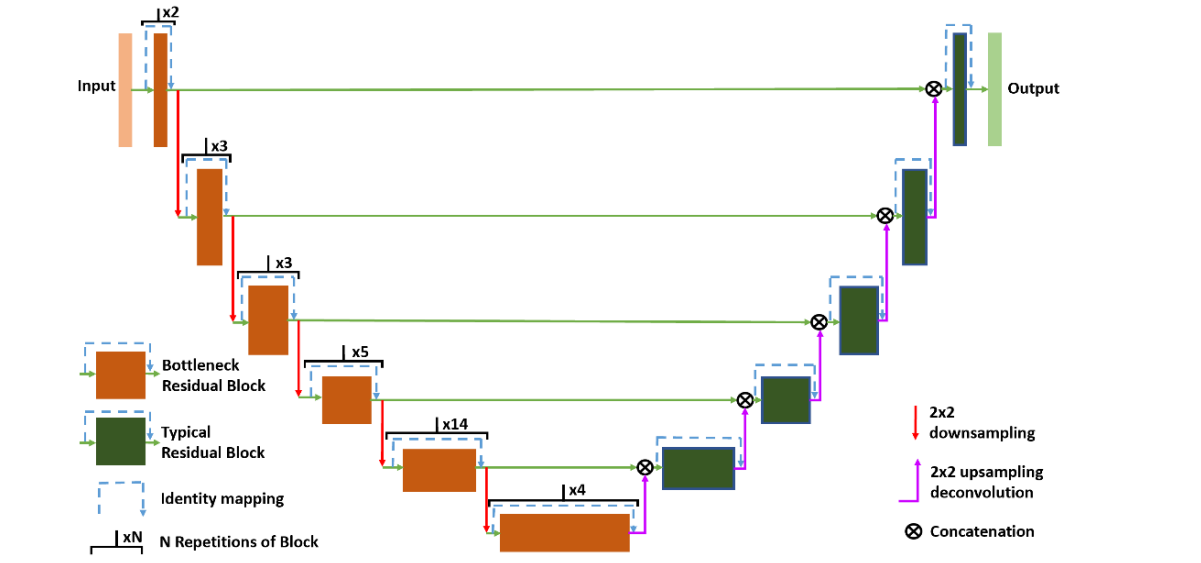
Brain tumor segmentation has become a crucial task in medical imaging, particularly in the analysis of MRI scans. Accurate segmentation of tumor regions is essential for diagnosis, treatment planning, and monitoring the progression of brain tumors. Over the years, various deep learning architectures have been employed to improve the accuracy and efficiency of tumor segmentation. These methods typically aim to identify and delineate tumor regions such as the whole tumor, tumor core, and enhancing tumor, which present unique challenges due to the complex nature of brain structures and tumor characteristics.

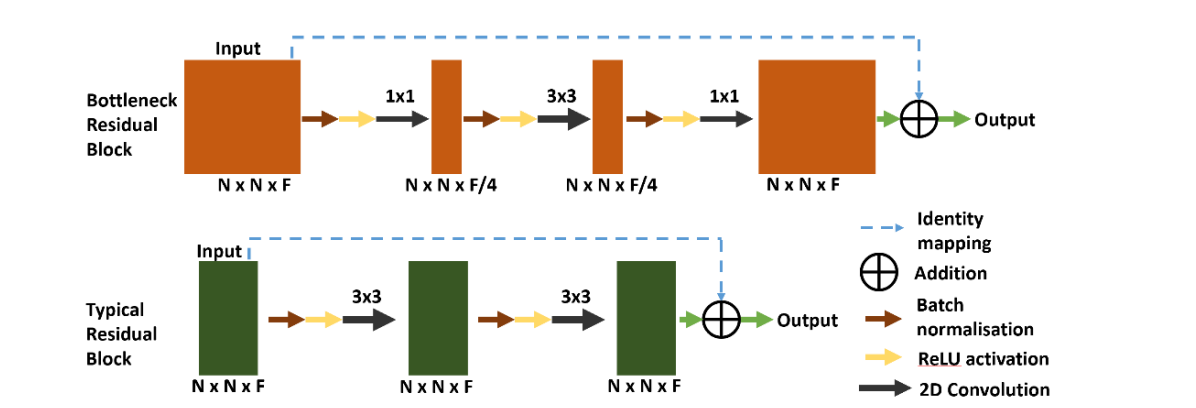
In recent studies, several authors have explored different deep learning models to address these challenges using the BraTS 2020 dataset, which provides a valuable resource for brain tumor segmentation. Below, we discuss the approaches and results of some notable works that have contributed to advancements in this field.

* In O. Çiçek et al., authors used a 3D U-Net model was used on BrasTS 2020 datset. The image dimensions of the input images were 128x128x128. The Dice Score of this model, mainly for Tumor core is 81.73%, for the Whole tumor is 90.09% and for the Enhancing tumor is 78.73%. These results display the success rate and functionality of 3D U-Net architectures for tumor segmentation in the human brain.

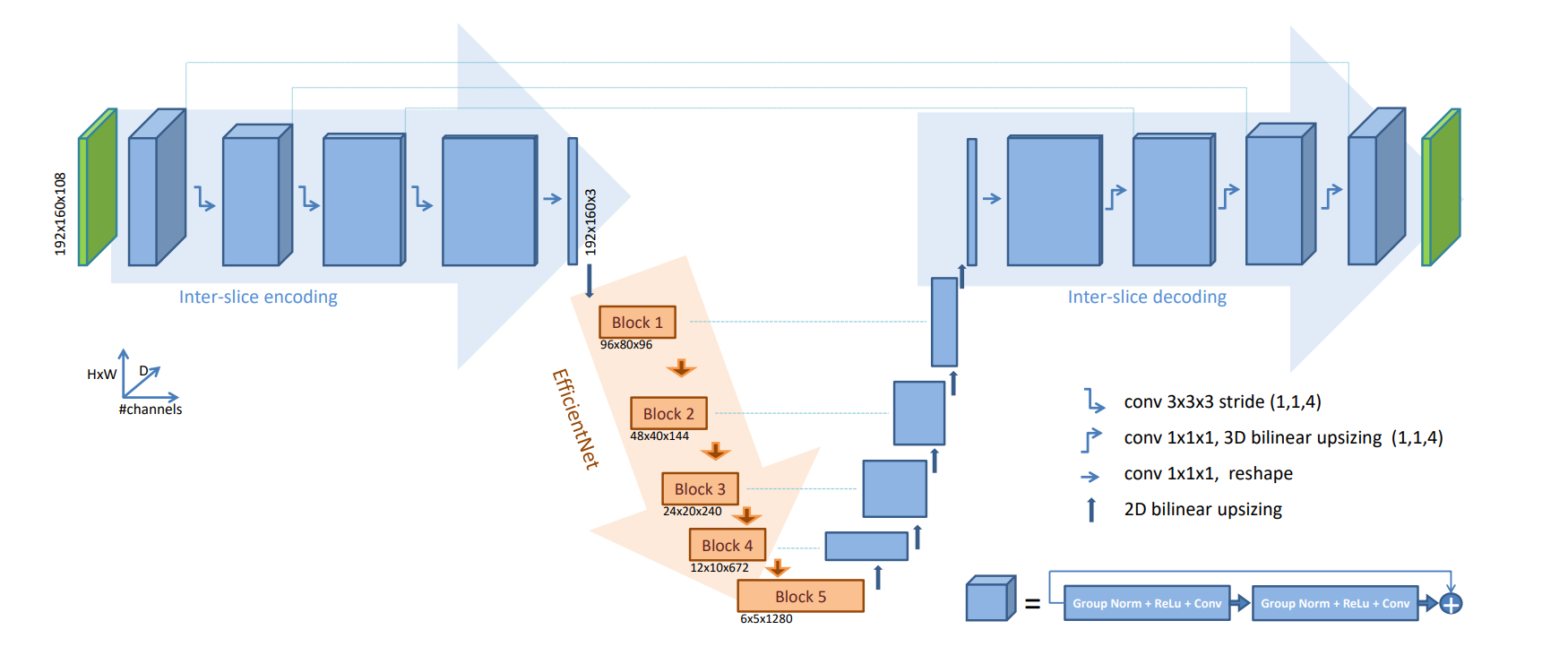
FIGURE 1

* In J. Colman et al., authors used a 2D Deep Residual U-Net model was used on BrasTS 2020 datset. The image dimensions of the 2D input image slices were 240x240. The Dice Score of this model, mainly for Tumor core is 79.83%, for the Whole tumor is 86.73% and for the Enhancing tumor is 75.14%. These results display the success rate and functionality of 2D Deep Residual U-Net architectures for tumor segmentation in the human brain.

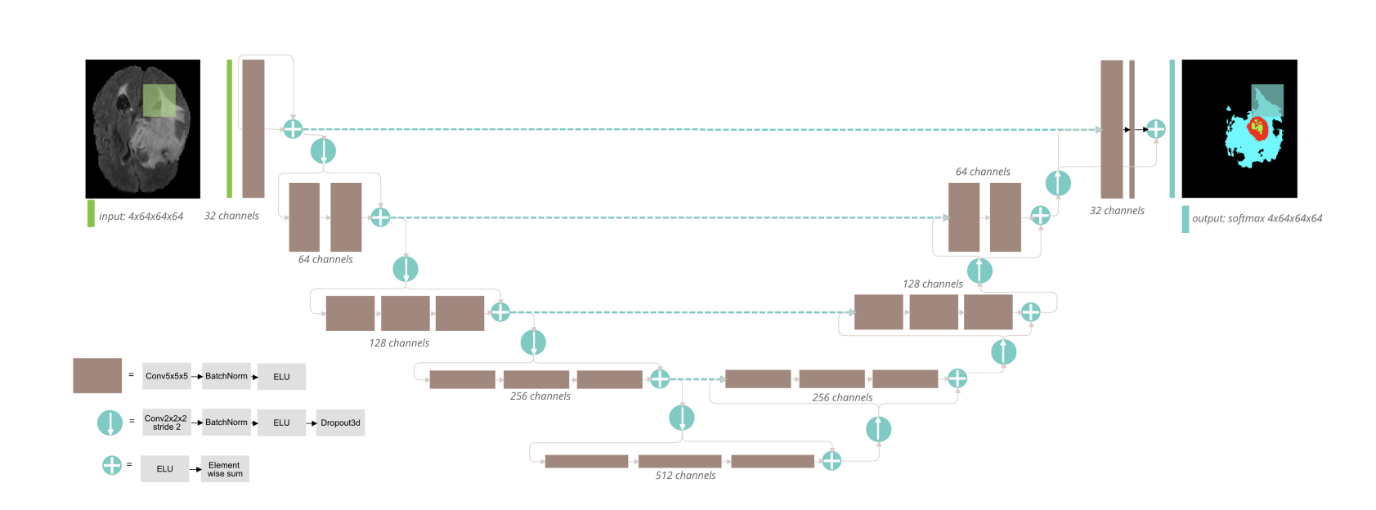




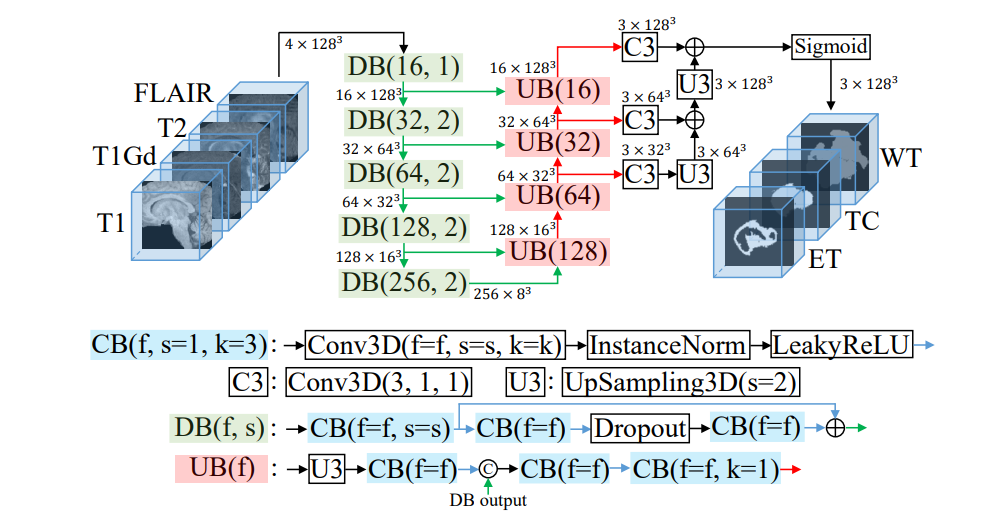
* In H. Messaoudi et al., authors used a 3D U-Net model was proposed on BrasTS 2020 datset, unlike the efficient usage of 2D classifier network model. The image dimensions of the input images were 192x160x108. The Dice Score of this model, mainly for Tumor core is 75.26%, for the Whole tumor is 80.68% and for the Enhancing tumor is 69.59%. These results display the success rate and functionality of re-utilizing the pre-existing 2D networks inside the 3D network architectures for improvizing the performance on tumor segmentation in the human brain.



* In L. M. Ballestar et al., authors used a 3D Encoder-Decoder-based V-Net model was used on BrasTS 2020 datset. The image dimensions of the input images were 64x64x64. The Dice Score of this model, mainly for Tumor core is 75.26%, for the Whole tumor is 84.63% and for the Enhancing tumor is 62.15%. These results display the success rate and functionality of 3D Encoder-Decoder-based V -Net architectures for tumor segmentation in the human brain. It is based on the goal of comparison and evaluation.



* In F. Wang et al, authors used a 3D U-Net model with it’s expertise in brain-wise patching stratergy was used on BrasTS 2020 datset. The image dimensions of the input images were 128x128x128. The Dice Score of this model, mainly for Tumor core is 79.8%, for the Whole tumor is 85.2% and for the Enhancing tumor is 77.8%. These results display the success rate and functionality of 3D U-Net architectures for improvised tumor segmentation in the human brain.



**Proposed Technique**

A ResUNet model was developed for addressing the image segmentation tasks. ResUNet is an effective and meaningful combination of U-Net and ResNet architectures for improvising the accuracy of the segmentation. The model is later named as ResUNet without Nearest Neighbour (Model –1). Nevertheless, the results were not up to the mark, thus introducing an advanced version, Modified Optimized ResUNet with Nearest Neighbour (Model –2). Both the versions are mainly focused on efficiency and segmentation.

**Model 1 - ResUNet without Nearest Neighbor:**

This is an encoder-decoder framework, in which the encoder extracts the hierarchical features with the help of ResNet blocks and the decoder rearranges the segmented output. The architecture consists of 4 channels (FLAIR, T1, T2, T1ce) and works only for input images of size 128x128

* **Input Layer:** The model’s input image is the size of 128x128 with 4 channels, contained by multiple MRI modalities.
* **Encoder:**
  + **ResNet Block 1:** The first block uses two 3x3 convolutions with residual connections and ReLU activation, outputting 64 feature maps.
  + **Max Pooling (Block 1):** Reduces the spatial dimensions from 128x128 to 64x64, capturing coarse features.
  + **ResNet Block 2:** The second block uses 128 filters, refining the features extracted in Block 1.
  + **Max Pooling (Block 2):** Reduces the spatial dimensions further to 32x32.
  + **ResNet Block 3:** Expands the feature depth to 256 filters for deeper feature extraction.
  + **Max Pooling (Block 3):** Further reduces the spatial dimensions to 16x16.
  + **ResNet Block 4:** The final ResNet block uses 512 filters, capturing high-level features.
  + **Max Pooling (Block 4):** The final downsampling step reduces the spatial dimensions to 8x8.
* **Bridge (Bottleneck):** At this stage, the model captures the most abstracted features with a ResNet block of 1024 filters.
* **Decoder (Output Layer):** A 1x1 convolution with softmax activation generates a 128x128 segmentation map, providing per-pixel classification into 4 distinct classes.

Even though, ResUNet model (without the Nearest Neighbour) displayed reasonable results, it could not satisfy the desired levels of Dice coefficients, mainly for precise segmentation.   
This takes place mainly due to the failing to maintain spatial fidelity in the upsampling and decoding phases. With this, lack of robust feature leads to further degrading of the sub-optimal segmentation accuracy.

Despite the fact that ResUNet has given better results compared to other models, its effectiveness in the segmentation is still a conversation.

**Model 2 - Optimized ResUNet with Nearest Neighbor:**

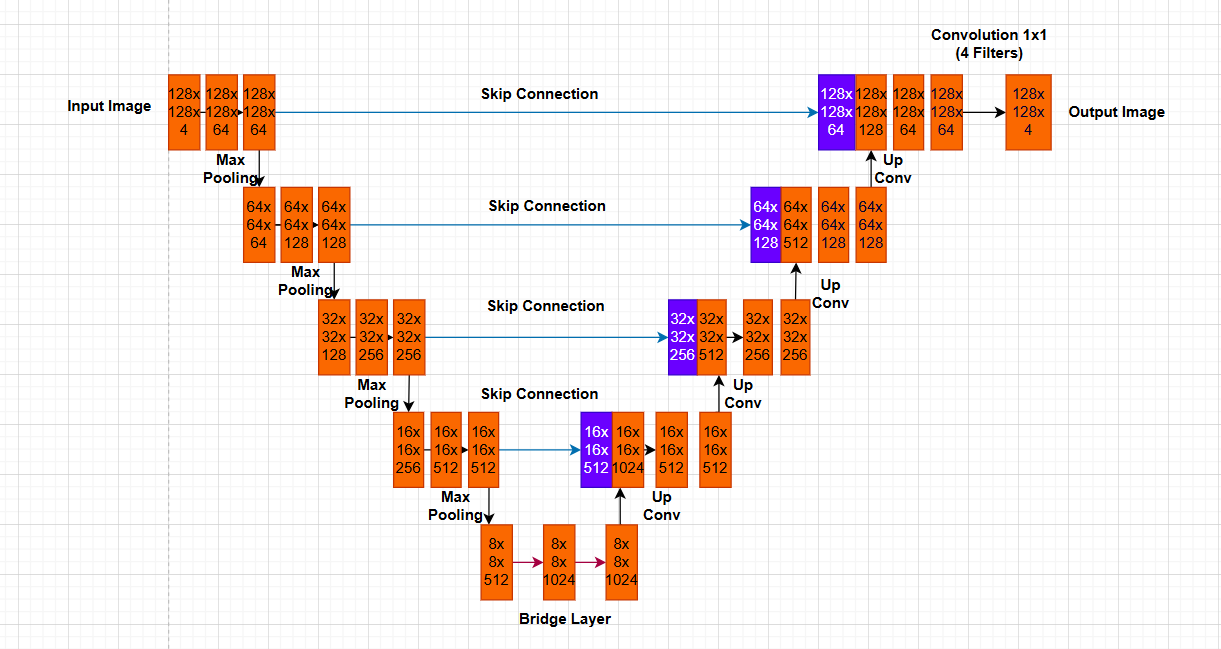
Optimized ResUNet with Nearest Neighbourwas built for addressing the limitations which were observed in the previous models. This is done by introducing nearest neighbour sampling inside the decoder for improving the feature refinement and increased segmentation accuracy.

* **Input Layer:** As with the first model, the input consists of a 128x128x4 image with 4 channels, including modalities such as **FLAIR, T1, T1ce, and T2**.
* **Encoder:**
  + **ResNet Block 1:** Outputs a feature map of 128x128x64, extracted through two 3x3 convolutions with 64 filters and residual connections.
  + **Max Pooling (Block 1):** Reduces the spatial dimensions to 64x64x64.
  + **ResNet Block 2:** Deepens the feature extraction with 128 filters, outputting a feature map of 64x64x128.
  + **Max Pooling (Block 2):** Reduces the spatial dimensions to 32x32x128.
  + **ResNet Block 3:** Further deepens the feature extraction with 256 filters, producing a feature map of 32x32x256.
  + **Max Pooling (Block 3):** Reduces the spatial dimensions to 16x16x256.
  + **ResNet Block 4:** Captures high-level features with 512 filters, producing a 16x16x512 feature map.
  + **Max Pooling (Block 4):** Reduces the spatial dimensions to 8x8x512.
* **Bridge (Bottleneck):** A ResNet block with 1024 filters extracts high-dimensional features from the bottleneck layer, maintaining rich feature representations.
* **Decoder:**
  + **Up Block 4:** Upsamples the bottleneck output to 16x16x512, concatenates it with the corresponding encoder feature map, and applies two 3x3 convolutions with 512 filters.
  + **Up Block 3:** Upsamples to 32x32x256, concatenates with the encoder feature map, and applies two 3x3 convolutions with 256 filters.
  + **Up Block 2**: Upsamples to 64x64x128, concatenates with the encoder feature map, and applies two 3x3 convolutions with 128 filters.
  + **Up Block 1:** Upsamples to 128x128x64, concatenates with the encoder feature map, and applies two 3x3 convolutions with 64 filters.
* **Output Layer:** A final 1x1 convolution with 4 filters and softmax activation generates a 128x128x4 output, with per-pixel class probabilities for the 4 classes.

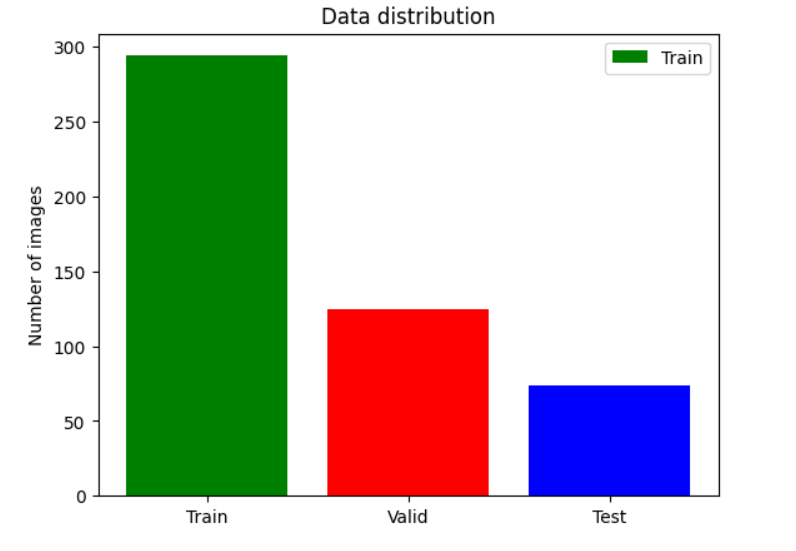
Model 2 is concluded as an better alternative as it addresses all the limitations in the model 1 by following the nearest neighbour upsampling and refining feature extraction. It is done in both the encoder and the decoder. Thus this model is concluded as a better alternative due to its better accuracy and better spatial resolution restoration , while being more precise at each pixel.

**Architecture of Our Best Performing Model(Model-2):**

**Modified ResUNet(Optimized) model**



**Spliting of Dataset(BraTS 2020):**

**Algorithm:**

**Step 1:** Start

**Step 2:** Initialize necessary libraries and configurations for the task.

**Step 3:** Load the dataset (Training and Validation).

**Step 4:** Perform data quantity analysis to determine how many folders and files are available in the dataset.

**Step 5:** Perform data visualization and display particular patient's MRI scan images to better understand the dataset.

**Step 6:** Define segmentation labels:

0: 'Not Tumor'

1: 'Necrotic/Core'

2: 'Edema'

3: 'Enhancing'

**Step 7:** Apply image slicing analysis to determine how many 2D slices are there in each 3D image.

**Step 8:** Split the training dataset in a ratio of 80:20 for training and testing, and use 100% of the validation dataset for validation analysis. Print the length of the training, testing, and validation datasets.

Number of training samples: **294**

Number of testing samples: **74**

Number of validation samples: **125**

**Step 9**: Visualize the data distribution.

**Step 10:** Perform data preprocessing using ImageDataGenerator with augmentation techniques like rotation, shifting, shearing, zooming, and flipping.

**Step 11**: Initialize a custom data generator (DataGenerator) class for handling data loading and augmentation.

* Use the ImageDataGenerator defined in **step 10** within the data generator.
* Define methods like \_\_len\_\_, \_\_getitem\_\_, on\_epoch\_end, and \_\_data\_generation to handle batch creation and data shuffling.

**Step 12:** Initialize individual generators for training, validation, and testing datasets.

training\_generator = DataGenerator(train\_ids)

valid\_generator = DataGenerator(val\_ids)

test\_generator = DataGenerator(test\_ids)

**Step 13:** Define the models for training:

**Model 1**: ResNet without nearest neighbor technique.

**Model 2:** Optimized ResNet with nearest neighbor technique.

**Step 14:** Train both models with different epochs/iterations (e.g., 15, 35, 50).

**Step 15:** Evaluate both models and take the best accuracy results from the two models with different epochs.

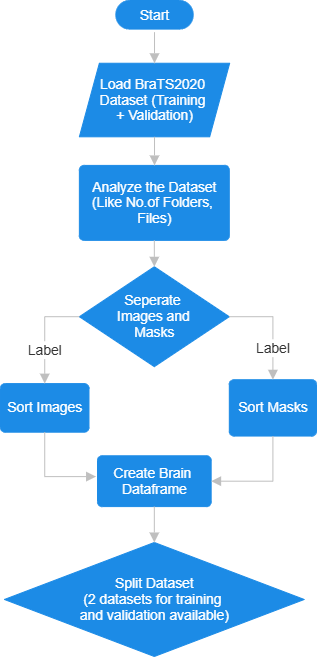
**Step 16:** Visualize the results of training and validation.

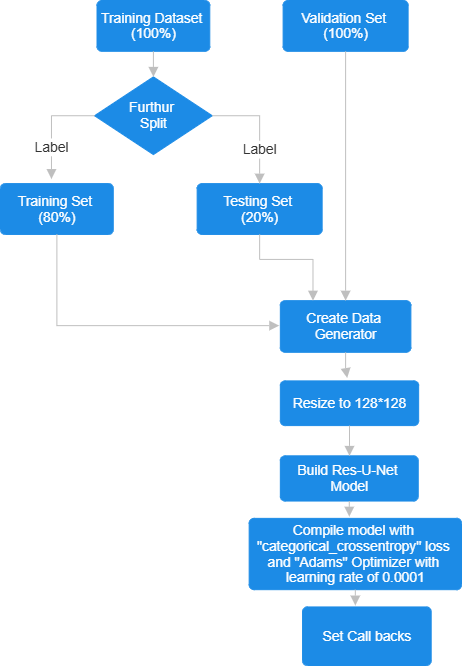
**Step 17:** Apply data segmentation techniques on the testing dataset and predict the results using the trained model.

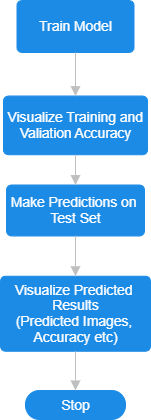
**Step 18**: Visualize the results of the predictions.

**Step 19**: Stop.

**Flow Chart**

****

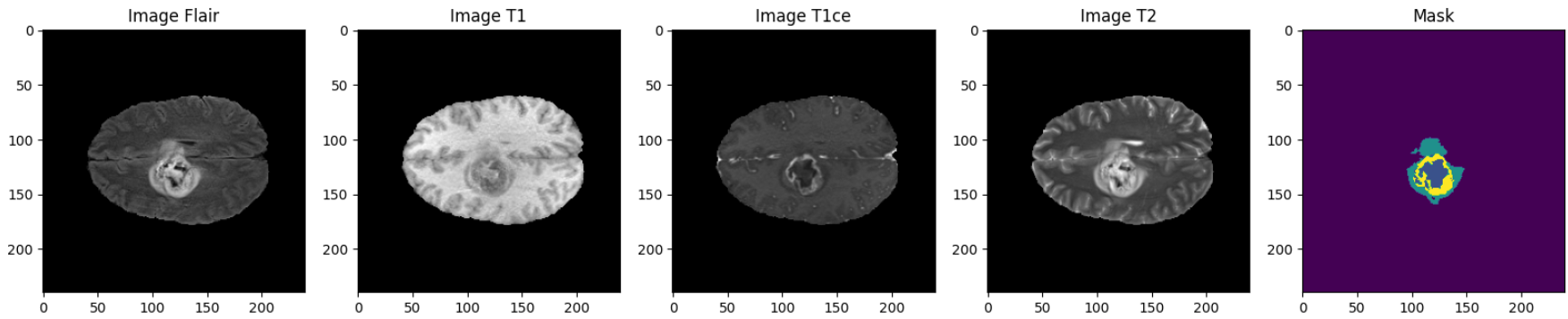
****

****

**Results:**

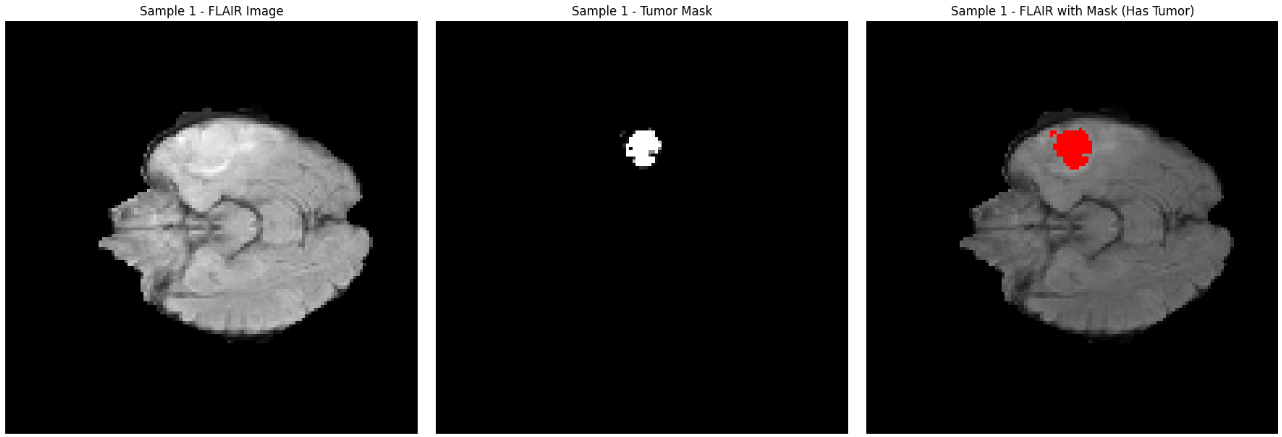
In this study, the selected dataset ‘BraTS 2020’ has played an important role by providing detailed and explained MRI images in different aspects which belong to different patients. These modalities showcase and understand different aspects of the brain’s pathology, providing an accurate result in the segmentation process of the tumor regions. The dataset contains different MRI sequences and is a combination of expert-level segmentation issues, which permits for only a robust analysis of the tumor and its boundaries, with the abnormalities attached with them. In this result section, we showcase the results and conclusions achieved through using the proposed models, emphasising the effectiveness of the appeared outcomes.

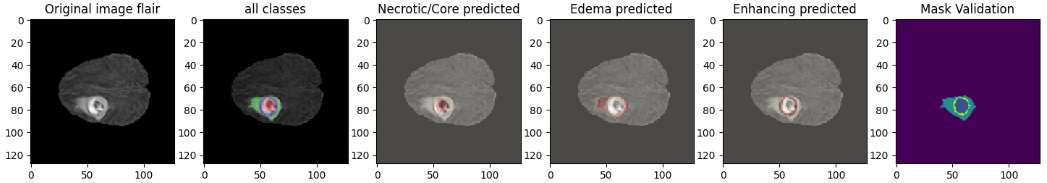
The dataset consists of various MRI images in different modalities. These modalities are arranged in different files according to the patients details and cases

****

The modalities present a detailed study and view of the tumor’s structure and tumor’s position with its abnormalities within the tumor, such as edema and swelling around the tumor or the fluid containing regions near it. The segmentation masks play a key role in representing different components with different colours, for enhancing the tumor core, the necrotic core and the peritumoral edema.

The complete dataset is divided into different batches, each specifically representing the respective modality. The code works on checking for the presence of tumor, by examining the non- zero values present inside the mask. An overlay is formed on the modality if a tumor is found in the specific region, therefore highlighting the part for displaying the proper difference between the normal and abnormal regions.



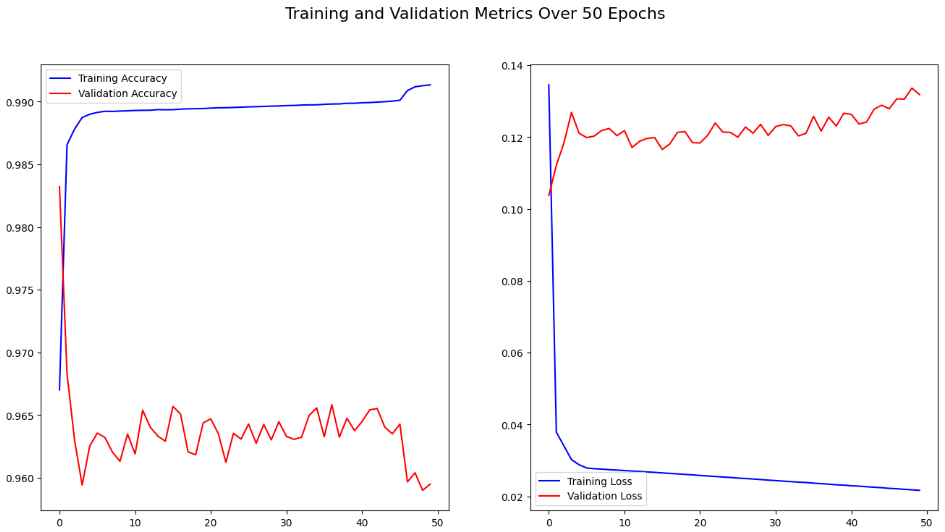


Both the models Model1 and Model2, are utilized for the predictions. The original image which contains all the details is given as an input. The tumor contained image consists of various classes and various parts, such as necrosis, edema and many more.

The loss metrics over 50 epochs, the validation accuracy and training are factors that present the performance of both the models 1 and 2. The accuracy levels of Model 1 is ~99% and the accuracy levels of Model 2 is ~99.5%. Nevertheless, fluctuation of the validation accuracy takes places, making the stabilizing levels of Model 1 around 96% and the Model 2 around 98%.

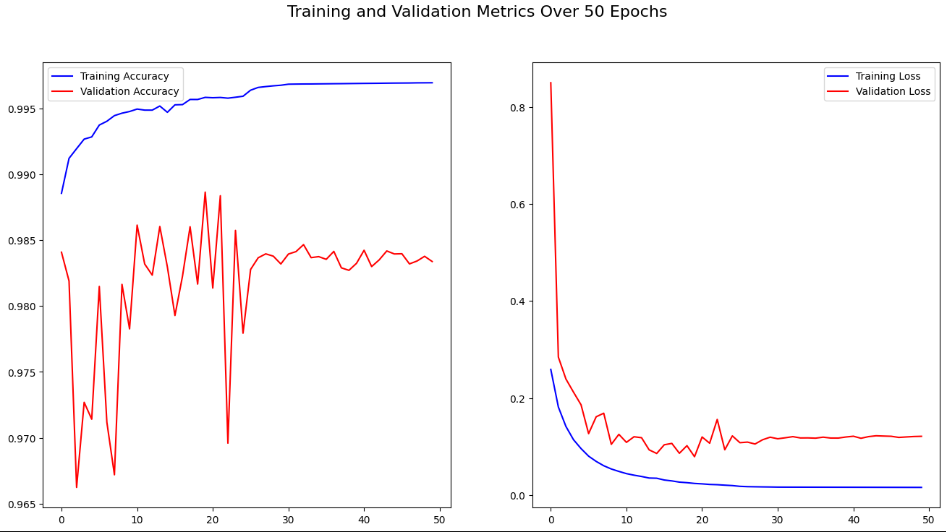
Even though both the models are created for the same purpose the training varies in both the models. In Model 1, there is a rapid decrease in the training loss, making it nearly zero. But in the Model 2 , the decrease takes place, but the final loss level is ~0.2, which makes Model 2 better than Model 1. Model 1 has a training loss of ~0.12.

**Model 1:**



* The training accuracy reaches around 99% and the validation accuracy fluctuates around 96%
* Even though the training loss decreases nearly to zero, the validation loss still remains around 0.12, represented and indicating overfitting.

**Model 2:**



* The training accuracy crosses the limit 99.5% and the validation accuracy fluctuates around 98%.
* Suggesting an overfitting issue with a little increased validation stability, from which the training loss drastically drops nearer to zero and the validation loss is stabilized

~0.2.

Below is the Comparison Table between training and validation metrics for the utilized model, which is ResUNet+.

|  |  |  |
| --- | --- | --- |
| Evaluation Metrics | Training Data | Validation Data |
| Accuracy | 0.99694 | 0.983376 |
| Dice Coefficient | 0.745992 | 0.504009 |
| Dice Coefficient (Edema) | 0.868952 | 0.428634 |
| Dice Coefficient  (Enhancing Tumor) | 0.782335 | 0.470367 |
| Dice Coefficient  (Necrotic Tumor) | 0.769888 | 0.373515 |
| Loss | 0.015743 | 0.121048 |
| Precision | 0.997855 | 0.984215 |
| Sensitivity | 0.997316 | 0.983382 |
| Specificity | 0.999287 | 0.99475 |

**Observations and Interpretations:**

* **Overfitting**: Even though both the models showcase the signs overfitting, mainly displaying a huge gap between the validation and training metrics. The validation loss of Model 1 mainly balances nearly at 0.12 and Model 2 nearly around 0.2. Both the models are highly expertized in training accuracy even though fluctuations take place in the issue of validation accuracy, which point the finger towards overfitting. The overfitting is more in Model 1 when compared Model 2, according to the clear observations.
* **Stability in Validation Metrics**: In the matter of stability Model 2 shows better stability levels in the loss over epochs and the validation accuracy in comparison to Model 1. The stability is displayed in the smaller fluctuations, thus proving that Model 2 is better than Model 1.
* **Comparative Performance**: Compared to Model 1, Model 2 displays to perform well , indicating the higher levels of validation accuracy.
* **Validation accuracy:** Even though there is an display of higher accuracy levels on the training data , the validation data is comparatively struggling to maintain the levels. Which suggests that Model 2 has better potential for overfitting and higher accuracy compred to Model 1.

**Conclusion:**

In the conclusion, Optimized ResNet with Nearest Neiahbor‘s (Model-1) performance is better compared to ResNet-U-Net (Model1) for the brain tumor segmentation. It is a marvellous option, by obtaining 98% of validation accuracy and 98.23% of testing accuracy. In comparison to Model-2, the Model-1’s validation accuracy is 96%. Even though, Res-U-Net+ contains the higher level of losses, it is comparatively more efficient and effective in finding and identifying tumor areas, that to when there are many types.

This study contains limitations, which include the specifications of using only the ‘BraTS’ dataset and include a training period 50 epochs. The model’s dependability and performance can be improved by training and testing the dataset for a longer period. And the future studies and improvisations must focus on the extending hand towards the training duration and experimenting with other architectures. For further improving of the model’s ability accuracy in finding the tumor types efficiently, exploring the new techniques such as usage of much larger and more variety of datasets is required.

Finally, this study showcases the role and importance of certain deep learning models in the brain tumor segmentation. This study, further answers many unsolved medical prophecies and complications regarding the brain tumor segmention in the medical field. This study’s main goal and aim is to diagnose patients at the right time, which can help in better treatment and patient care. The future updating work on the refining of the model, which leads this to clinical and medical field, in the real-world.