# Medical Image Segmentation

**[1]Sania Shaikh, [2]Khushi Singh, [3]Aishwarya Nagpure, [4]Arpit Mohankar, [5]Swati Bhatt**

[1]Undergraduate Student, [2]Undergraduate Student, [3]Undergraduate Student, [4]Undergraduate Student, [5]Assistant Professor
Department of Artificial Intelligence and Data Science,
Rajiv Gandhi Institute of Technology, Mumbai, India
[1]sania.shaikh9202@gmail.com, [2]khushimangal2003@gmail.com, [3]aishwarya8403@gmail.com,
[4]arpitmohankar24@gmail.com, [5]swatibhatt238@gmail.com

***Abstract*— Medical image segmentation in medical imaging plays a crucial role in helping clinicians accurately identify abnormal areas in magnetic resonance images. MRI sequences are essential for distinguishing tumors by analyzing the contrast and texture of soft tissues, making accurate segmentation. This paper focuses on presenting deep learning architectures for the automated segmentation of abnormal regions in MRI scans, with a focus on brain tumors. It incorporates ResNet50 architecture to classify the presence of a tumor, while ResUNet, VGG19-UNet, and UNet models focus on precise segmentation. The dataset used comes from The Cancer Imaging Archive (TCIA), which features MRI scans from 110 patients with lower-grade gliomas and includes manual segmentation masks for fluid-attenuated inversion recovery (FLAIR) abnormalities. Key performance metrics such as accuracy for classification and Tversky loss, Dice coefficient, and IoU for segmentation ensure the effective identification of tumor regions.**

*Index Terms*— **Brain tumor segmentation, Magnetic Resonance Imaging (MRI), ResNet50, ResUNet, UNet, VGG19 UNet.**

## I. INTRODUCTION

Medical image segmentation is a crucial task in healthcare, allowing for accurate identification and outlining of anatomical structures and abnormal regions in imaging techniques such as MRI, CT scans, and X-rays. Accurate segmentation is essential for various clinical applications, including disease diagnosis, treatment planning, and surgical guidance. Traditionally, medical image segmentation relied mainly on rule-based algorithms or manual annotations, which were not only time-intensive but also varied between observers. However, with the emergence of deep learning architectures such as convolutional neural networks (CNNs), this process has become significantly more precise, faster, and highly automated, delivering reliable pixel-wise classifications that surpass traditional approaches. Additionally, advancements in both computational power and model architecture have further enhanced medical image segmentation. Models like U-Net, Convolutional Networks, and SegNet are popular because of their ability to capture detailed as well as global information. These models follow an encoder-decoder architecture, with the encoder capturing essential features and the decoder enhancing the segmentation results. They have demonstrated high performance. Despite these advancements, the field still faces significant challenges. A key issue is the limited availability of annotated medical datasets, which are vital for training the models. Furthermore, variations in imaging methods, acquisition protocols, and patient demographics make it difficult for models to generalize effectively. To address these challenges, researchers are focusing on data-efficient training strategies such as semi-supervised learning, data augmentation, and transfer learning. Additionally, efforts are underway to improve model robustness and ensure consistent performance across various clinical settings, which is crucial for broader adoption in real-world healthcare environments.

## II. RELATED WORK

The paper [1] presents three architectures for brain MRI analysis. The first model classifies MRI images into affected and unaffected regions using 75x75 patches from 256x256 images. It features 10 convolutional and 4 max-pooling layers and achieves 99.63% accuracy but requires a separate decoder for segmentation. The second architecture, an autoencoder, uses skip connections for better spatial retention, achieving 98.84% segmentation accuracy, and a Dice score of 91.76%. The third model incorporates an upsampling attention module, improving feature extraction with pixel- and channel-level attention, achieving 99.49% segmentation accuracy and a 99.81% F-score.

The ResUNet+ model workflow starts with preprocessing MRI images by normalizing multiple modalities (T1, T2, and FLAIR) to extract the Region of Interest (ROI). It then utilizes a hybrid architecture combining residual blocks and attention mechanisms within a U-Net framework to enhance feature extraction and reduce semantic gaps. Trained and evaluated on datasets like BraTS 2020, 2019, and 2018, the model outperforms state-of-the-art methods [2].

The proposed system [3] enhances MRI brain images by converting them to 2D grayscale, reducing noise with median filtering, and improving quality through adaptive histogram equalization. Data augmentation addresses resolution variations, while segmentation isolates the skull and outlines the tumor using OTSU-based thresholding and active contours. Key features are extracted for classification, and a modified CNN achieves 95.6% accuracy. Advanced segmentation further refines tumor detection and density estimation using Gaussian kernel distribution.

The preprocessing phase focuses on denoising MRI brain scans using wavelet-based thresholding, including Haar, Symlet, Morlet, and Daubechies wavelets. These methods effectively remove noise while preserving key image features like edges. Performance is evaluated using SNR, PSNR, and MSE, with Daubechies level 3 and Haar wavelets yielding the best results. Haar and Symlet wavelets, due to their smaller support, are particularly effective in capturing fine details [4].

The digital image processing technique begins with MRI images as input, which undergo preprocessing to enhance size and shape, followed by segmentation using the Chan-Vese algorithm. It is followed by feature extraction using principal component analysis (PCA) and gray level co-occurrence matrix (GLCM) techniques to improve accuracy. The images are classified using deep convolutional neural networks (DCNN) [5].

The authors [6] present a methodology involving skull stripping for brain cortex extraction, followed by intensity-based segmentation of cerebrospinal fluid, gray matter, and white matter. Tumor regions are identified using region-based algorithms

analyzing pixel area properties. Extracted features like mean, entropy, energy, and variance are classified with a feedforward neural network. Performance evaluation through training and testing metrics demonstrates improved brain tumor detection accuracy.

SegNet is a deep neural network designed for semantic segmentation that is pixel-wise, featuring an encoder-decoder structure based on VGG16. Its decoder upsamples feature maps using pooling indices from the encoder, eliminating the need for learned upsampling. Compared to models like FCN and DeepLab, SegNet balances memory efficiency with segmentation accuracy [7].

Swin-Unet leverages Swin Transformer blocks for segmentation. The encoder splits images into patches, creating sequence embeddings, while the decoder uses patch expansion and skip connections to merge features, preserving spatial information. This architecture enhances segmentation accuracy through efficient feature representation and fusion [8].

This paper [9] examines two medical image segmentation models: a 12-layer transformer-based encoder (ViT) and a hybrid encoder (R50-ViT) combining ResNet-50 with ViT. Both use ImageNet pretraining with a 224x224 input resolution. TransUNet, evaluated against U-Net and AttnUNet, performs best due to its ability to capture both global and local details. Hybrid architectures (R50-ViT) outperform pure transformer-based models, with skip connections, resolution, and patch size fine-tuned for optimal results.

## III. APPROACH

### *System Architecture*

Figure 1 shows the system architecture. The proposed system has two main components, the tumor classification, and the segmentation component. First, the classifier checks whether the MRI contains a tumor. If the model predicts no tumor with high confidence, the image is skipped, and no further segmentation is performed. If the classifier detects a tumor, the MRI is passed to the ResUNet, VGG19-UNet, and UNet segmentation models to predict the location of the tumor.

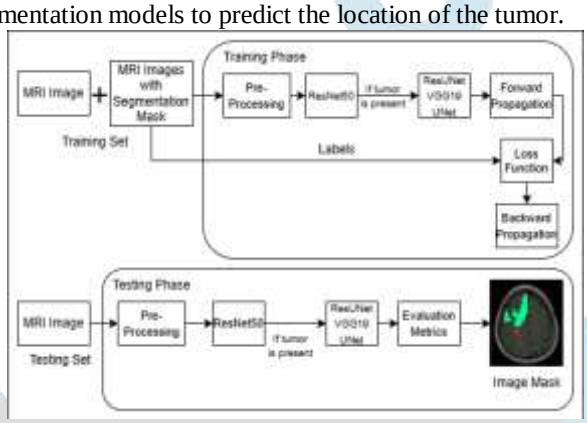


Fig. 1. System Architecture

### *Dataset*

The LGG Segmentation Dataset comprises MRI scans of lower-grade gliomas, complete with manual segmentation masks and tumor regions. The dataset includes brain MRI images in the FLAIR (Fluid-Attenuated Inversion Recovery) sequence, sourced from The Cancer Imaging Archive (TCIA) and featuring data from 293 patients. The FLAIR sequence suppresses signals from fluids (like cerebrospinal fluid), making it easier to identify lesions or abnormalities related to disease. In these MRIs, hyperintense (bright) areas may indicate abnormal tissues such as edema, demyelination, or tumors, distinguishing them from normal brain structures. The dataset is split into 85% for training and validation and 15% for testing, with the former being further divided into 90% for training and 10% for validation.

### *Classification*

1) ResNet50: The MRI images and their corresponding segmentation masks are loaded and preprocessed. Pixel values are normalized by dividing by 255 to fit within a range of [0, 1]. The images are resized to 256x256 pixels, and a grid displaying the images with their segmentation masks is created, with the masks overlaying the images at partial transparency for better visualization. A pre-trained ResNet50 model, excluding the top layers, serves as the backbone for classification, allowing the model to utilize previously learned features from extensive datasets. The ResNet50 foundation model is followed by fully connected layers, beginning with an average pooling layer to minimize spatial dimensions and avoid overfitting. For input into fully linked layers, a flatten layer transforms 2D feature maps into a 1D vector. After that comes a dropout layer that randomly deactivates neurons during training to prevent overfitting, as well as a thick layer with 256 units and ReLU activation. A second dense layer with 256 units and ReLU activation follows, this time with dropout for regularization. Two units with softmax activation make up the last dense layer, which allows binary classification to determine whether or not a tumor is visible in the picture. For this classification job, the loss function is categorical crossentropy, and the performance parameter utilized for training and validation is accuracy. A model checkpoint saves the top-performing model based on validation loss, and an early stopping function stops training if the validation loss does not improve for 15 consecutive epochs, preventing overfitting.

### *Segmentation*

1) VGG19-UNet: MRI scan data and corresponding segmentation masks are processed by reading metadata from a CSV file containing image and mask file paths. The paths are sorted and mapped to patient IDs, ensuring alignment. Masks are analyzed to classify the presence (1) or absence (0) of tumors. Images and masks are resized to 256x256 resolution, normalized for consistent intensity, and expanded to maintain compatibility with the model. Batch processing and data shuffling improve training efficiency and reduce overfitting.

The segmentation model follows a U-Net architecture with a VGG19 encoder pretrained on ImageNet. The encoder extracts hierarchical features through four convolutional blocks, progressively reducing spatial dimensions from 256x256x3 to 32x32x1024. A bottleneck layer (16x16x1024) connects the encoder to the decoder, which reconstructs spatial dimensions through transposed convolutions and skip connections. The decoder upscales the feature maps back to 256x256x64, with a final 1x1 convolution producing single-channel masks. A sigmoid activation function generates pixel-wise probabilities for tumor segmentation. The model

consists of 28 layers and approximately 20 million parameters. It uses the Adam optimizer with a dynamic learning rate scheduler and Tversky loss function, which helps handle class imbalance. Regularization techniques such as dropout and batch normalization are applied. With a batch size of 16, the model is trained over 200 epochs, and early stopping is used to avoid overfitting. Tversky loss is used to assess performance, and the anticipated and ground truth masks are visually compared. Figure 2 shows MRI with original and predicted masks.

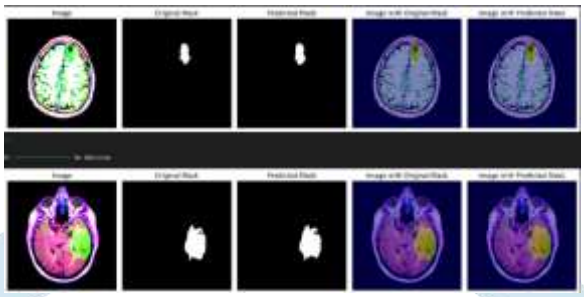

Fig. 2. Segmentation Masks of VGG19-UNet Model

2) UNet: The dataset consists of MRI images and their corresponding segmentation masks, with metadata stored in a CSV file. Missing values in the CSV are filled using the most frequent value. Each MRI scan is mapped to its respective mask, and metadata is integrated with image and mask file paths. Images are converted into tensors, and mask values are normalized to binary (0 or 1). Data augmentation techniques, including random brightness/contrast adjustments, channel dropout, and color jitter, enhance training diversity. The data is then batched and shuffled to improve training efficiency.

Nine essential phases are formed by the network's four encoder blocks, bottleneck layer, four decoder blocks, and output block. There are two convolutional layers per block, for a total of 18 convolutional layers. Three-channel pictures (such as RGB) are converted into 64 feature maps by the input layer using a double convolution with a 3×3 kernel, stride of 1, and padding of 1. Each convolution is followed by batch normalization, and ReLU activation aids in the preservation of spatial information. Using max-pooling layers, downsampling blocks gradually decrease the spatial size while deepening the features. The input is changed from 64 to 512 channels via these blocks. In order to preserve spatial information, upsampling blocks concatenate encoder features via skip connections and use transposed convolutions to restore the spatial resolution. The feature depth is decreased from 1024 to 64 channels during the upsampling steps.

A sigmoid activation function ensures that the expected mask values stay between 0 and 1 after a final 1×1 convolution layer transfers the feature maps to a single-channel output for binary segmentation. Adam is used to optimize the model after it has been trained using Binary Cross-Entropy (BCE) loss. Gradients are calculated for weight updates over several epochs during training, but no gradient calculations are made during validation. For analysis, the loss values are recorded. After training, the model is evaluated using unobserved data, producing binary masks by thresholding predictions at 0.5. The anticipated masks are superimposed on ground truth photos to display the outcomes. Metrics like the Dice coefficient and Intersection over Union (IoU) are used to measure segmentation accuracy and performance. Figure 3 shows the output for the UNet model.

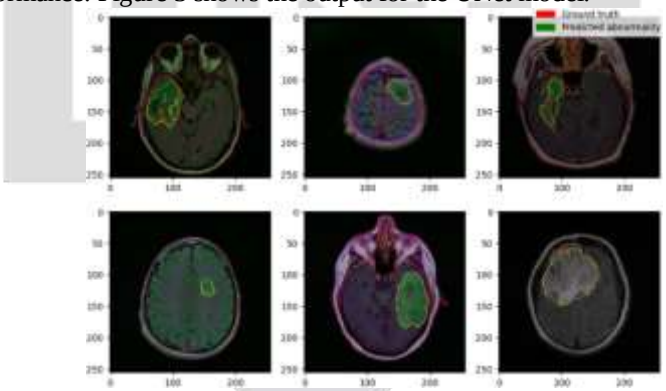

Fig. 3. Segmentation Masks UNet Model

3) ResUNet: If the classifier detects a tumor, the MRI image is forwarded to the ResUNet segmentation model to locate the tumor. This architecture combines the U-Net structure with residual blocks to capture both spatial and feature information. The model processes an input image of size (256, 256, 3) through convolutional layers that integrate residual blocks, which enhance feature extraction while facilitating better gradient flow and deeper model performance. Each residual block consists of two convolutional layers, batch normalization, and ReLU activation, with a shortcut connection that bypasses the convolutional layers to maintain essential features. Through a series of convolutional layers and pooling operations, the downsampling route (encoder) gradually reduces the size of the input image while deepening the feature maps. From 16 in stage 1 to 256 in stage 5 (the bottleneck), there are more filters in the downsampling stages. Following this phase, the model starts upsampling in order to return the feature maps to their initial size. In order to recover finer details lost during downsampling, skip connections are used to merge low-level and high-level features from the decoder with feature maps from earlier encoder stages during upsampling. The feature maps are refined using a residual block following each upsampling step. With Upsample 1 merging with conv4 (128 filters) and Upsample 4 merging with conv1 (16 filters), the upsampling stages are identical to the downsampling stages. Finally, a binary mask for tumor segmentation is produced via a 1x1 convolutional layer with sigmoid activation. This mask is a single-channel (grayscale) image, and the presence of a tumor is indicated by pixel values close to 1, and its absence is indicated by values close to 0. A callback is used to end training early if validation loss does not improve after 20 epochs in order to avoid overfitting. MRI with both the original and predicted segmentation masks is shown in Fig. 4.

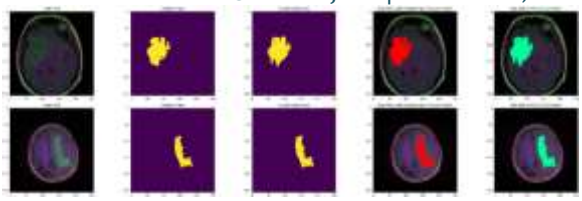

Fig. 4. Segmentation Masks of ResUNet Model

## IV. RESULTS

MRI segmentation is implemented using U-Net, ResUNet, and VGG19-UNet models. These architectures are employed to enhance segmentation accuracy, and their performance is evaluated based on various metrics. A web application is created using Streamlit, demonstrating the effectiveness of these models in delineating MRI structures, contributing to a robust and user-friendly website. Figure 5 shows the webpage where an MRI is given as input, and the original and the predicted masks are displayed along with the pixel count.

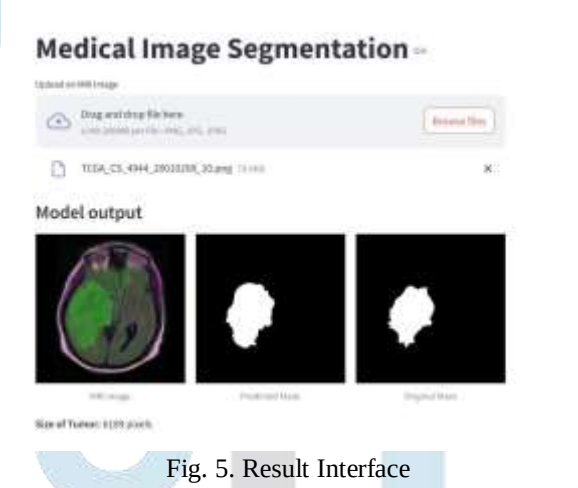


Fig. 5. Result Interface

Figure 6 shows an alert box stating that there is no tumor in case of its absence.


Fig. 6. Result Interface in Absence of Tumor

***Evaluation Parameters***

Evaluation metrics are utilized to assess the performance of the partial implementation. The evaluation metric used for classification is accuracy, while Dice Coefficient, IoU, and Tversky Loss are used for segmentation.

1) Accuracy: The accuracy of the classification model is determined using the following formula:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Where, TP = true positive, TN = true negative, FP = false positive, FN = false negative.

2) Dice Coefficient: The Dice Coefficient measures the similarity between two sets (e.g., predicted and ground truth segmentation masks). It is defined as:

$$Dice = \frac{2 \times |A \cap B|}{|A|+|B|} \quad (2)$$

Where, A is the set of pixels in the predicted mask, B is the set of pixels in the ground truth mask, and (A∩B) is the number of overlapping pixels (true positives). The range is [0,1], where 0 means no overlap and 1 means perfect overlap.

3) Intersection over Union (IoU): IoU measures the ratio of the intersection of two sets to their union:

$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (3)$$

The range is [0,1], where 0 means no overlap and 1 means perfect overlap.

4) Tversky Loss: The Tversky loss is designed to handle class imbalance by introducing two parameters, α and β, that control the weight given to false positives and false negatives, respectively. It is especially beneficial for tasks where minimizing either false positives or false negatives is a priority.

$$\text{Tversky Loss} = 1 - \frac{|A \cap B|}{|A \cap B| + \alpha|A \setminus B| + \beta|B \setminus A|} \quad (4)$$

Where: (A∩B) is the intersection of the predicted and ground truth masks (true positives).
A\B the difference between the predicted mask and the ground truth mask, indicating false positives.
B\A is the difference between the ground truth mask and the predicted mask (false negatives).
α controls the penalty for false positives, and β controls the penalty for false negatives.
α and β provide control over the level of emphasis placed on false positives and false negatives, respectively.

Table 1 shows the results for the different architectures used.

Table I Results

| *Classification* | |
|---|---|
| ResNet50 | Accuracy: 88.93% |
| *Segmentation* | |
| VGG19-UNet | Tversky Score: 89.57% |
| UNet | Dice Coefficient: 0.8571<br>IoU Score 0.7500 |
| ResUNet | Tversky Score: 89.98% |

The classification model offers an accuracy of 88.93%. Among the segmentation models, UNet is evaluated using the Dice coefficient (0.85) and IoU score (0.75), while ResUNet and VGG19 UNet are assessed using the Tversky score, achieving 89.98% and 89.57%, respectively.

***Discussions***

Figure 7 shows the classification model loss, which demonstrates strong generalization, with training and validation accuracy converging after 15 epochs and loss approaching zero, indicating no overfitting. The final test accuracy is 88.93%. The model excels in detecting tumors (high recall of 0.94 for Class 1) but has lower precision (0.79), leading to more false positives. Conversely, for Class 0 (no tumor), precision is high (0.96), but recall is lower (0.86). The confusion matrix shows 330 true negatives, 52 false positives, 13 false negatives, and 193 true positives. To improve precision for tumor detection, model tuning could focus on reducing false positives through decision threshold adjustments or class weighting.

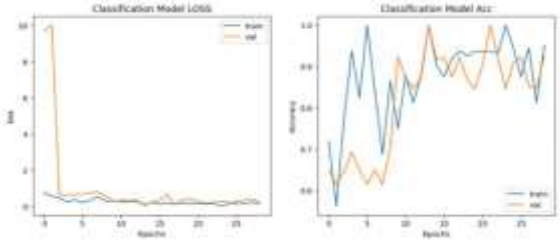

Fig. 7. Classification Model Loss

The graphs in Fig. 8 illustrate the training performance of the VGG19 UNet segmentation model using the focal Tversky loss function. Effective learning and little overfitting are indicated by the loss graph's consistent drop in both training and validation loss. The training and validation curves closely follow one another, indicating high generalization, while the Tversky score graph shows a steady growth. Strong model performance is confirmed by the final evaluation, which shows a loss of 0.1607 and a Tversky score of 91.21% with a final segmentation accuracy of 89.57%. Overall, the model learns and generalizes effectively for segmentation tasks, while slight variations in early validation results point to some instability.

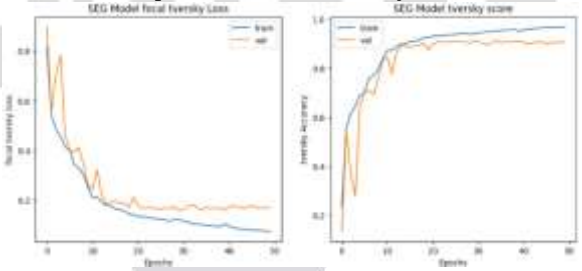

Fig. 8. VGG19-UNet Model Loss

The UNet model's training vs. validation loss graph is displayed in Fig. 9. Effective learning is seen by the graph's sharp decline in training and validation loss during the first few epochs. Validation loss varies little, indicating mild overfitting, although training loss keeps decreasing steadily. The overall loss of 0.012 is minimal in spite of these fluctuations, suggesting strong model performance. Although strategies like early halting or dropout could further increase stability, the near alignment of validation and training loss trends points to reasonable applicability.

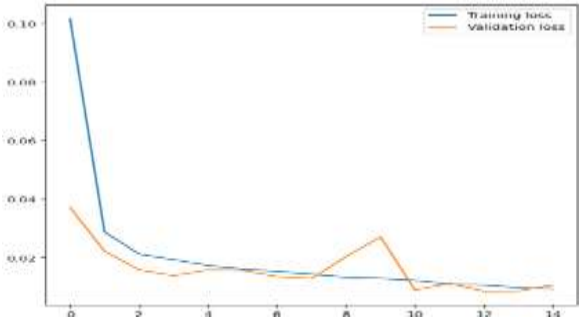


Fig. 9. UNet Model Loss

The performance of the segmentation model is assessed using Tversky loss over 60 epochs, as shown in Fig. 10. While the validation loss varies and stabilizes at epoch 15, suggesting minimal progress for validation data, the training loss gradually declines, demonstrating good learning. Although it is not severe, the difference between training and validation losses points to modest overfitting. Both the training and validation scores in the Tversky score plot increase rapidly over the course of 15 epochs before

plateauing, with the training score continuously above the validation score, suggesting possible overfitting once more. Although generalization might be strengthened, the validation score, which stabilizes at about 0.89, indicates good segmentation performance.

Overall, the model achieves high accuracy but shows signs of slight overfitting. Early stopping or regularization may help improve generalization. The final Tversky score on test data is 89.98%.

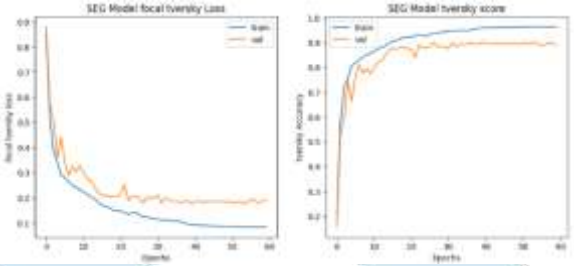

Fig. 10. ResUNet Model Loss

## V. CONCLUSION

In conclusion, the project utilized the LGG Segmentation Dataset of MRI scans to develop a comprehensive model for detecting and segmenting lower-grade gliomas. By employing a ResNet50 architecture for classification and ResUNet, UNet, VGG19 models for segmentation, the approach effectively leveraged advanced deep learning techniques. The data preprocessing steps included augmentations to enhance model generalization, and both classification and segmentation tasks were evaluated using robust metrics. The model achieved a high test accuracy for classification, with strong performance in identifying tumors, although some signs of overfitting were noted. Overall, while the model demonstrated promising results, further tuning and regularization could help improve its generalization capabilities, ultimately enhancing its utility in clinical settings for brain tumor analysis.

## REFERENCES


[1] S. Subramaniam, K. B. Jayanthi, C. Rajasekaran, and R. Kuchelar, "Deep Learning Architectures for Medical Image Segmentation," *IEEE 33rd Int. Symp. Computer Based Medical Systems (CBMS)*, Rochester, MN, USA, pp. 579-584, 2020.

[2] S. Metlek and H. Çetıner, "ResUNet+: A New Convolutional and Attention Block-Based Approach for Brain Tumor Segmentation," in *IEEE Access*, vol. 11, pp. 69884-69902, 2023.

[3] A. R. P. Sinha, M. Suresh, N. Mohan R., A. D., and A. G. Singerji, "Brain Tumour Detection Using Deep Learning," *Seventh Int. Conf. Bio Signals, Images, and Instrumentation (ICBSII)*, Chennai, India, pp. 1-5, 2021.

[4] M. Gurbină, M. Lascu, and D. Lascu, "Tumor Detection and Classification of MRI Brain Image using Different Wavelet Transforms and Support Vector Machines," *42nd Int. Conf. Telecommunications and Signal Processing (TSP)*, Budapest, Hungary, pp. 505-508, 2019.

[5] S. Somasundaram and R. Gobinath, "Early Brain Tumour Prediction using an Enhancement Feature Extraction Technique and Deep Neural Networks," *Int. J. Innovative Technology and Exploring Engineering (IJITEE)*, vol. 8, no. 10S, pp. 170-174, 2019.

[6] S. Damodharan and D. Raghavan, "Combining Tissue Segmentation and Neural Network for Brain Tumor Detection," *Int. Arab J. Information Technology*, vol. 12, no. 1, pp. 42-52, 2015.

[7] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder Decoder Architecture for Image Segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481-2495, 2017.

[8] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, "TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation," *arXiv preprint arXiv:2102.04306*, 2021, in press.

[9] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, "SwinUnet: Unet-like Pure Transformer for Medical Image Segmentation," *European Conf. Computer Vision*, pp. 205-218, 2022.