

Brain Tumor Segmentation using Deep Learning-Based MRI Analysis

Ankit Kumar Mandusia

Department of Physics and Computer Science
Dayalbagh Education Institute
Agra, India

Prof. C. Vasantha Lakshmi

Department of Physics and Computer Science
Dayalbagh Education Institute
Agra, India

II. RELATED WORKS

Abstract— Brain tumors represent a complex and challenging group of neoplasms ranking among the deadliest cancers globally, with gliomas being the predominant type of primary brain tumor. Cancerous glial cells in the brain and spinal cord give rise to these tumors. Advancements in diagnosing and treating them are crucial. Detecting brain tumors early is challenging for medical professionals due to tumors' multiplicity and complex appearance in medical imaging. This study utilizes, a deep learning approach for brain tumor segmentation from different types of complex magnetic resonance imaging (MRI) scans is presented. An innovative method that combines Grad-Cam [1] (Gradient-weighted Class Activation Mapping) with an existing deep neural network for brain tumor segmentation is proposed. The MRI images are preprocessed to enhance contrast and normalize intensity levels. The method is trained on a combination of large datasets of brain scans, fine-tuning its weights to optimize the segmentation performance obtained.

Keywords— Gradient-weighted Class Activation Mapping, EfficientNet, Convolutional Neural Network.

I. INTRODUCTION

Brain tumors are the most common cancers in the world and it is one of the critical problems for clinicians to segment brain tumors on medical images for correct diagnosis, crafting treatment plans, and ongoing disease surveillance. The challenges faced by those with brain tumors can vary widely, as some tumors may grow swiftly, leading to an array of issues ranging from minor to severe. This research has the intention to integrate Grad-Cam [1] enhancement with existing deep learning models such as EfficientNetB0, EfficientNetB7, U-Net[7], and a CNN architecture in this task. For the process to be maximally effective and systematic a three-stage structure is applied for conducting the study. The first phase addresses the preprocessing stage encompassing the image cropping and image enhancement processes. In the second stage, there are several neural network architectures, which have been learned above, are implemented. Finally, Grad-Cam [1] is integrated with the network's last layer. This step is crucial as It allows us to see which image regions hold the most weight for the neural network's predictions, thus providing interpretability and reinforcing trust in automated decisions.

Daimary et al.[2] investigated the use of hybrid convolutional neural networks (CNNs) to address the challenges of automatic segmentation with high accuracy. Their research explored combining established CNN architectures like SegNet, U-Net, and ResNet-18 to create U-SegNet, Res-SegNet, and Seg-UNet . They found that these hybrid models achieved superior performance (mean accuracy of 91.6%, 93.3%, and 93.1% respectively for U-SegNet, Res-SegNet, and Seg-UNet) compared to traditional methods by mitigating the loss of information from small tumors during downsampling steps through skip connections.

Zheng et al.[3] proposed an enhanced U-Net architecture to address shortcomings in brain tumor segmentation. Their method utilizes a serial encoding-decoding structure with hybrid dilated convolutions and a new loss function that focuses on difficult samples, with the mean Dice coefficient improving to 86.9% and the Hausdorff distance decreasing to 25.79 mm. Despite these improvements, issues like handling intricate tumor edges and optimizing loss functions persist, indicating room for further refinement. Limitations of commonly used loss functions, like binary cross-entropy and Dice loss, were not completely addressed.

Chandan Yogananda et al.[4] introduced an innovative method on a small dataset for segmenting gliomas in their Study. They utilized a 3D-Dense-UNet architecture which consisted of three separate networks. The evaluation results demonstrated the algorithm's effectiveness, as it achieved a mean Dice score of 0.92 across cross-validation folds. This study highlights the potential of their approach in accurately identifying and segmenting gliomas, which can greatly aid in the diagnosis and treatment planning process for brain tumors. One notable limitation is the requirement for a significant number of subjects for network training, which is a common challenge in this field. Also, the algorithm showed a slight decrease in performance when applied to a clinical dataset, indicating its potential sensitivity to real-world variations in imaging parameters.

Michal Futrega et al.[5], proposed research aimed at improving U-Net structure for brain tumor division in the BraTS21 contest, researchers explored various techniques including deep supervision loss, Focal loss, decoder attention, drop block, and residual connections. Their approach employed a thorough removal study to recognize the most efficient model setups, resulting in impressive achievement in the contest—emerging victorious in the validation stage and securing third spot in the test phase.

The findings emphasized the original U-Net as the most efficient model, attaining an average Dice score of 0.9130 across five folds. Among the different options explored, deep supervision significantly boosted performance slightly (average Dice score of 0.9149). Yet, despite these breakthroughs, the research accepted substantial obstacles in manual division processes, which are laborious and suffer from disparities due to the varied shapes and looks of the brain. The study also highlighted the added computational requirement brought about by some of the more intricate model expansions, like the variational autoencoder branch in the SegResNetVAE model, which significantly extended training times without a corresponding enhancement in accuracy. Baiju Babu Vimala et al [6], investigated brain tumor classification using a transfer learning approach with EfficientNets to classify brain tumors into three categories: glioma, meningioma, and pituitary tumors. Five pre-trained models from the EfficientNets family, ranging from EfficientNetB0 to EfficientNetB4, are fine-tuned using the publicly accessible CE-MRI Figshare dataset. The study revealed that optimizing the EfficientNetB2 model led to notable enhancements in performance which achieved an impressive overall accuracy of 99.06% on the test set, with precision, recall, and F1-scores all surpassing 98%. The study recognizes the difficulty of limited dataset size, especially in accurately classifying meningioma, suggesting the importance of larger datasets to improve model reliability.

III. METHODOLOGY

The objective of this work is to integrate the deep CNN models for the segmentation of tumors in the brain MRI images. The following flowchart provides a comprehensive overview of the process, illustrating the sequence and interaction of each phase.

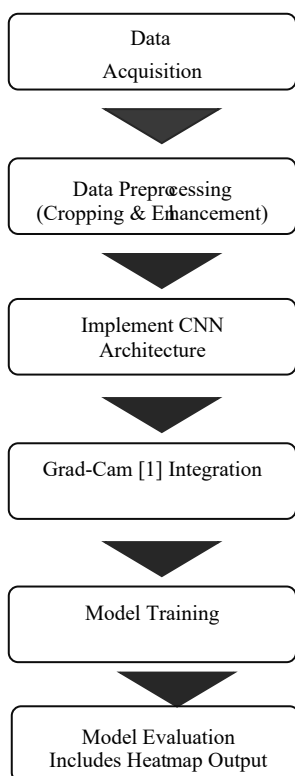


Fig. 1 Overview of Study

A. Data Preprocessing

In this stage, data preprocessing and data augmentation are performed. Image cropping plays a pivotal role, especially when focusing on regions of interest where brain tumors are located. This targeted approach not only enhances the relevance of the analysis but also optimizes computational efficiency and accuracy.

Steps Involved in Image Cropping:

- **Grayscale Conversion:** Convert the input color image into grayscale format (RGB to Gray). Grayscale images contain only intensity information and are frequently more manageable in image processing tasks.
- **Gaussian Blur:** After converting to grayscale, Gaussian blur is applied for smoothing operation that helps reduce noise and fine details. This step improves the image quality for subsequent processing.
- **Binary Thresholding:** In the thresholding step, we set a threshold value of '45'. Pixels in the grayscale image with intensity values above this threshold are converted to white while those below are turned to black (zero).
- **Contour Detection:** After Binary Thresholding extract contours from the binary thresholded image. Contours are simply the boundaries of connected regions with similar pixel intensity.
- **Extreme Point Identification:** Identifies the extreme points (left, right, top, and bottom) of the largest contour. These extreme points are determined by finding the lowest and highest x and y coordinates of the contour.

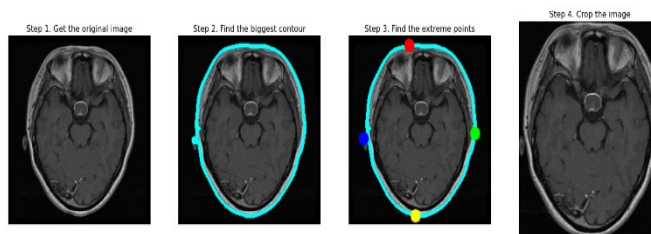


Fig. 2 Image Cropping Process

The integration of these preprocessing procedures successfully isolates the region of interest (i.e., the brain tumor) from the original image.

Data augmentation stands as the most commonly utilized technique to enrich the quality and diversity of a training dataset. In this task, images are randomly rotated within ± 10 degrees and subjected to brightness adjustments with scaling factors ranging from 0.85 to 1.15. Spatial transformations include translations up to 0.2% of image dimensions and shearing by 12.5 degrees. Zoom adjustments are made as necessary, with no zoom represented by a factor of 0. Horizontal flipping is utilized to increase data variability, while vertical flipping is not applied. Finally, any empty areas resulting from these transformations are filled using the nearest pixel values.

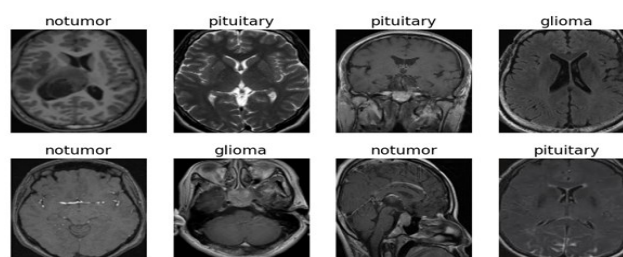


Fig. 3 Augmented Images from preprocessed dataset

B. Implement CNN Architecture

Various CNN architectures are evaluated and implemented, each offering unique advantages and suited for different aspects of the task. Below, the specifics of each architecture—U-Net, a standard CNN, EfficientNetB7, and EfficientNetB7—are discussed, explaining their roles and configurations within the research.

For the Standard CNN model, a standard CNN architecture designed to efficiently process 150x150 pixel RGB images is utilized. The architecture begins with a primary convolutional layer that employs 32 filters of size 4x4 to detect initial features. This is followed by a max pooling layer with a 3x3 window to minimize the spatial dimensions. Subsequently, the network progresses through layers with 64 filters, coupled with another max pooling phase to further refine the detected features. For deeper pattern recognition the network employs a convolutional layer with 128 filters. The last layers of the network are as follows: a flattening process; a dense layer with 512 nodes and a dropout mechanism in order to avoid overfitting; and lastly, a softmax layer for classification into four categories.

The U-Net[7] model is utilized for sophisticated segmentation of 150x150 RGB images and consists of two primary sections: an encoder that extracts features and a decoder that ensures precise localization. The encoder begins with a convolutional layer using 32 tiny 3x3 filters activated by ReLU to identify basic patterns, followed by a 2x2 max pooling layer to reduce the image size. The complexity of feature extraction is increased by doubling the filters to 64, still using 3x3 kernels and ReLU activation, before another 2x2 max pooling step further diminishes the dimensions. In the decoder section, the advanced features from the encoder's final stage are upscaled and merged with the initial simple features, blending detailed local data with more generalized information. This is refined by a convolutional layer with 32 filters of 3x3 size and ReLU activation. The network culminates with a 1x1 convolution that has a filter count equivalent to the number of target classes, which is four.

The EfficientNet B7 architecture is chosen from the EfficientNet Family suite for its depth and complexity, the B7 model excels in extracting features with high precision from 150x150 RGB images. It's a model that scales effectively, balancing network dimensions and resolution through a unique coefficient. Starting with a foundation of well-optimized convolutional layers for initial feature detection, it progresses to more intricate layers adept at discerning subtle data differences.

Conversely, the EfficientNetB0 variant is utilized for its swift processing and accurate performance, making it ideal for the preliminary evaluation of 150x150 RGB images. As the smallest in the series, B0 retains the network's original proportions while scaling efficiently, catering to scenarios demanding fast turnaround and low computational demand. Despite a resolution lower than B1 and B2, B0's robust feature detection and classification capabilities render it suitable for early image analysis and instantaneous processing.

C. Grad-Cam [1] Integration

Grad-Cam [1], standing for Gradient-weighted Class Activation Mapping, serves as a visualization tool in deep learning. Its unique attribute lies in its architectural agility and adaptability. In this research, Grad-Cam [1] is implemented alongside various CNN models to fulfill segmentation objectives, showcasing its capability to conform to different model structures while providing insightful visualizations.

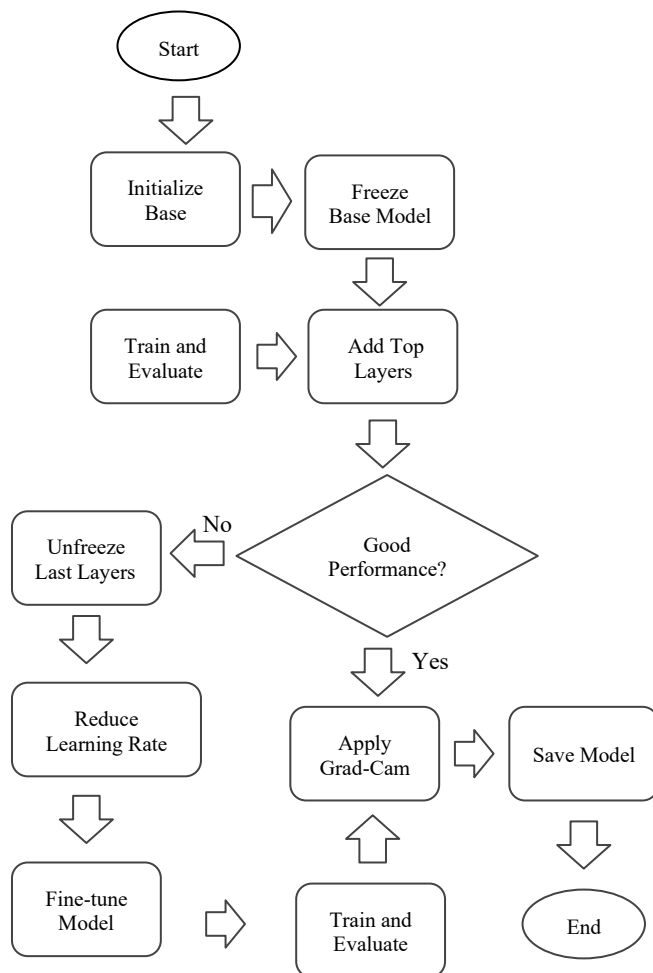


Fig. 4 Architecture of Integration of Grad-Cam [1] with CNN Model

The process for generating a segmentation heatmap with Grad-Cam [1] involves several critical steps:

- **Forward Pass:** The input image (150 x 150) is fed into the trained CNN or Pre-Trained CNN, resulting in the final class scores and the activation maps of a chosen convolutional layer (typically the last one before the final dense layers).

- **Gradient Calculation:** Backpropagation is performed to compute the gradients of the class score of interest with respect to the activation maps of the last convolutional layer.
- **Weighted Activation Map:** Multiply each feature map by its corresponding gradient value and then sum up the weighted feature maps to obtain a single heatmap, effectively spotlighting key areas of influence.
- **Heatmap Generation:** The resulting weighted activation map is normalized and then overlay the heatmap on the original input image to visualize the regions. This heatmap visually represents the model's focus for the predicted class.

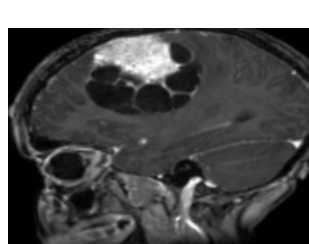


Fig. 7 Input Image

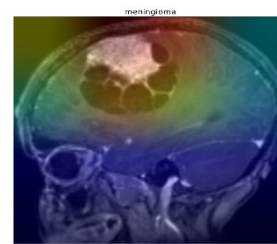


Fig. 8 Output Image

IV. RESULTS AND DISCUSSION

A. Dataset Description

The dataset utilized in this study combines three different sources: "Figshare," "SARTAJ dataset," and "BrT35H." By combined datasets, we get a total of 7,023 MRI images of human brains, which have been categorized into four groups: "glioma," "meningioma," "no tumor," and "pituitary." Specifically, the images in the "no tumor" category are exclusively obtained from the "Br35H" dataset and later combined with other sources.

B. Evaluation Metrics

To assess the model's effectiveness, four different types of parameters, which are accuracy, loss, Intersection over Union (IoU), and Dice coefficient are used. These metrics give us useful info, but they focus on different aspects of the model's performance.

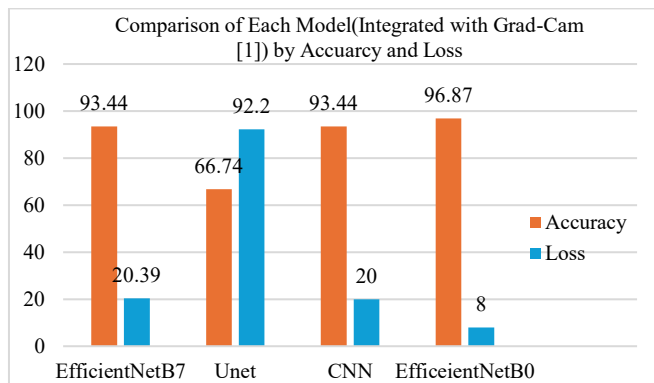


Fig. 5 Comparison of Models

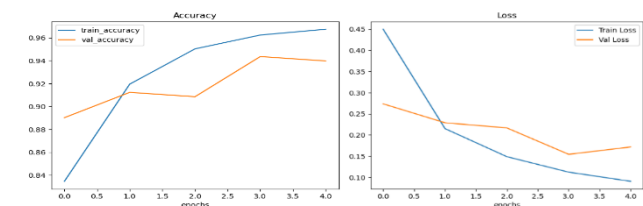


Fig. 6 Loss and Accuracy of EfficientNetB0

After completing the initial epochs, we noted the accuracy and loss among all neural network architectures we used for the task, the convolutional CNN architecture achieved an

accuracy of 93.44% and a loss of 20%, While using Grad-Cam [1] with the EfficientNetB0 model which is pre-trained model, we achieved a higher accuracy of 96.87% and a loss drop near 8%. This suggests that EfficientNetB0 is more effective in generating segmentation outputs that closely align with the actual tumor regions.

C. Results

This study delves into tumor localization within medical scans, likely centered on the brain. It uses a method known as Grad-Cam [1] to produce visual overlays on the scans that spotlight potential tumor locations. These overlays help reveal the characteristics that the models deem significant for identifying tumors. By analyzing these heatmaps, we can extract insights into the features the models prioritize for accurate tumor classification.

The top-performing models were from the EfficientNet series, with EfficientNetB0 reaching an IoU of 98.28% and a Dice coefficient of 98.57%. The EfficientNetB7 was also notable, with an IoU of 96.33% and a Dice coefficient of 96.90%.

A conventional Convolutional Neural Network (CNN) also showed strong results, matching the 96.33% IoU and 96.90% Dice coefficient of the EfficientNetB7.

The U-Net model had less favorable outcomes, with an IoU of approximately 78% and a Dice coefficient near 80%.

Table 1: Performance of all Model

Model (Integrated With Grad-Cam [1])	Global Accuracy	Mean Dice	Mean IOU
EfficientNetB7	96.45	97.56	95.28
Unet	66.74	80.23	78.49
CNN	93.44	96.90	96.32
EfficientNetB0	96.87	98.50	96.33

It is observed that Employing Grad-Cam [1] enhanced our understanding of how these models process and interpret medical scans. The EfficientNetB0 model is notably effective, achieving high scores in accuracy, IOU, and Dice Coefficient metrics, demonstrating its potential usefulness in clinical environments. Although the U-Net architecture showed lower performance, it highlights the need for further development. Both EfficientNetB7 and the standard CNN models also performed well. This research advances the application of deep learning in medical diagnostics and paves the way for future studies that might focus on refining these models or expanding the datasets used.

Table 2: Performance Comparison with Existing Studies

Paper	Model Name	Dataset	Dice Score
Michal Futrega (2021)	UNet	BraTS21	0.9130(Mean)
Ping Zheng(2022)	Improved U-Net	Kaggle Open Source (2764 images)	0.9262(Mean)
MD Abdullah (2023)	2d- UNet	BraTS 2017,2018, 2019 and 2020	0.95
Dinthisrang Daimary (2020)	CNN	BraTS 2012	0.93124
Jakhongir Nodirov(2022)	3D U-Net	BraTS 2020	0.8974
Ramin Ranjbarzadeh (2021)	Cascade CNN	BRATS 2018	0.9014
Proposed Model	Grad-Cam [1] with EfficientNetB0	SARTAJ, FigShare, BrT35H	0.9850

Among the evaluated models, our model, which combines Grad-Cam [1] with EfficientNetB0, reached an impressive Dice score of 0.9857. This reduces overfitting and improves the model's accuracy on unseen data. Our approach not only demonstrates improved performance metrics but also sets a new benchmark in leveraging large datasets for training more effective medical imaging models.

V. CONCLUSION

This experiment investigated the development of a deep learning-based system for automatic brain tumor segmentation within 3D magnetic resonance imaging (MRI) scans. Our primary objective was to enhance clinical decision-making in brain tumor diagnosis and treatment planning. By combining Gradient-weighted Class Activation Mapping (Grad-Cam [1]) with pre-trained deep learning models, we aimed to create a flexible and effective tool that could handle various tumor types and MRI modalities. which was trained on a large and diversely sourced dataset. By employing extensive preprocessing methods, including cropping of the MRI scans to focus on relevant features, which helped in reducing noise and improving the model's focus on tumor regions.

Our study investigated the combination of Grad-Cam [1] with multiple neural network frameworks, including EfficientNetB0, EfficientNetB7, U-Net, and a custom CNN architecture. The results demonstrate that EfficientNetB0 achieved superior performance. In contrast, both the custom CNN architecture and EfficientNetB7 demonstrated relatively lower performance in all evaluated metrics, while the U-Net architecture resulted in the lowest performance.

REFERENCES

- [1] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra, "Grad-Cam: Visual Explanations from Deep Networks via Gradient-Based Localization," 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, pp. 618-626, doi: 10.1109/ICCV.2017.74.
- [2] D. Daimary, M. B. Bora, K. Amitab, and D. Kandar, "Brain Tumor Segmentation from MRI Images using Hybrid Convolutional Neural Networks," Procedia Computer Science, vol. 167, pp. 2419-2428, Jan. 2020, doi: 10.1016/j.procs.2020.03.295.
- [3] P. Zheng, X. Zhu, and W. Guo, "Brain tumor segmentation based on an improved U-Net," BMC Medical Imaging, vol. 22, no. 1, Nov. 2022, doi: 10.1186/s12880-022-00931-1.
- [4] C. G. B. Yogananda et al., "A fully automated deep learning network for brain tumor segmentation," Tomography, vol. 6, no. 2, pp. 186-193, Jun. 2020, doi: 10.18383/j.tom.2019.00026.
- [5] M. Futrega, A. Milesi, M. Marcinkiewicz, and P. Ribalta, "Optimized U-Net for brain tumor segmentation," in Lecture notes in computer science, 2022, pp. 15-29. doi: 10.1007/978-3-031-09002-8_2.
- [6] B. Babu Vimala, S. Srinivasan, S. K. Mathivanan, et al., "Detection and classification of brain tumor using hybrid deep learning models," Sci Rep, vol. 13, Art. no. 23029, 2023. doi: 10.1038/s41598-023-50505-6.
- [7] O. Ronneberger, P. Fischer, and T. Brox, "U-NET: Convolutional Networks for Biomedical Image Segmentation," in *Lecture notes in computer science*, 2015, pp. 234-241. doi: 10.1007/978-3-319-24574-4_28.
- [8] H. Mohan and K. Chatterjee, "Detection of brain abnormality by a novel Lu-Net deep neural CNN model from MR images," Machine Learning With Applications, vol. 2, p. 100004, Dec. 2020, doi: 10.1016/j.mlwa.2020.100004.
- [9] Rintu Joseph and Mr. Sanoj C Chacko, "Brain Tumor Detection & Classification using Machine Learning" In ICCIDT-2023 (Volume 11, Issue 01) (pp. ICCIDT2K23-301). doi:10.17577/ICCIDT2K23-301.G.
- [10] S. Chatterjee, F. A. Nizamani, A. Nürnbergger, and O. Speck, "Classification of brain tumors in MR images using deep spatio-spatial models," Scientific Reports, vol. 12, no. 1, Jan. 2022, doi: 10.1038/s41598-022-05572-6.
- [11] Narayanan Krishnasamy and Thangaraj Ponnusamy, "Deep learning - based robust hybrid approaches for brain tumor classification in magnetic resonance images," International Journal of Imaging Systems and Technology, Oct. 2023, doi: <https://doi.org/10.1002/ima.22974>.
- [12] "Early Detection of Brain Tumor and Survival Prediction Using Deep Learning and An Ensemble Learning from Radiomics Images | IEEE Conference Publication | IEEE Xplore," [ieeexplore.ieee.org. https://ieeexplore.ieee.org/document/9971932](https://ieeexplore.ieee.org/https://ieeexplore.ieee.org/document/9971932).
- [13] J. Nodirov, A. B. Abdusalomov, and T. K. Whangbo, "Attention 3D U-Net with Multiple Skip Connections for Segmentation of Brain Tumor Images," Sensors (Basel), vol. 22, no. 17, Aug. 2022, Art. no. 6501. doi: 10.3390/s22176501
- [14] R. Ranjbarzadeh, A. Bagherian Kasgari, S. Jafarzadeh Ghouschi, et al., "Brain tumor segmentation based on deep learning and an attention mechanism using MRI multi-modalities brain images," Sci Rep, vol. 11, Art. no. 10930, 2021. doi: 10.1038/s41598-021-90428- 8.