Iranian Journal of Medical Physics

ijmp.mums.ac.ir

Deep Learning-Based Brain Tumor Segmentation in MRI Images: A MobileNetV2-DeepLabV3+ Approach

Farhad Abedinzadeh Torghabeh[1], Seyyed Abed Hosseini[2*]

1. Department of Biomedical Engineering, Mashhad Branch, Islamic Azad University, Mashhad, Iran.
2. Department of Electrical Engineering, Mashhad Branch, Islamic Azad University, Mashhad, Iran.

ARTICLE INFO

Article type:
Original Paper

Article history:
Received: July 25, 2023
Accepted: Dec 04, 2023

Keywords:
Brain Tumor
Magnetic Resonance Imaging
Artificial Intelligence
Segmentation Stage

ABSTRACT

Introduction: Brain tumors (BTs) pose significant challenges in medical diagnosis and treatment due to their heterogeneity and complex characteristics. Accurate and precise segmentation of BTs in magnetic resonance images (MRIs) is crucial for effective treatment planning and patient care. In this study, we propose an ensemble deep-learning (DL) model to address the challenging task of BT segmentation. We aim to achieve accurate localization and delineation of tumor regions across different axial views.

Material and Methods: The dataset used in this study consists of 3064 T_1-weighted contrast-enhanced MRI images obtained from patients diagnosed with glioma, meningioma, and pituitary tumors. Image preprocessing techniques, including normalization and intensity transformation, were applied to enhance the contrast and standardize the intensity values. The DL model is based on the DeepLabV3+ architecture combined with three well-known deep convolutional neural networks as encoders: MobileNetV2, ResNet50, and XceptionNet.

Results: The proposed ensemble model, with MobileNetV2 as the encoder, demonstrated superior performance in BT segmentation. The model achieved an average dice similarity coefficient of 0.938 and a global accuracy of 0.997. Compared to alternative models, MobileNetV2-DeepLabV3+ showed significant accuracy and segmentation precision improvements.

Conclusion: The ensemble DL model, leveraging the strengths of MobileNetV2 and DeepLabV3+, offers a robust and efficient solution for accurate BT segmentation in MRI images. The model's ability to delineate tumor regions holds great promise for enhancing diagnosis and treatment planning in BT analysis. Future work will explore further fine-tuning techniques and evaluate the model's performance on larger datasets to assess its generalization capabilities.

►*Please cite this article as:*
Abedinzadeh Torghabeh F, Hosseini SA. Deep Learning-Based Brain Tumor Segmentation in MRI Images: A MobileNetV2-DeepLabV3+ Approach. Iran J Med Phys 2024; 21: 343-354. 0.22038/ijmp.2023.73972.2313.

Introduction

Brain tumors (BTs) are characterized by the aberrant proliferation of cellular structures within the cranial cavity or its adjacent structures. These neoplastic growths originate from diverse cell populations present within the brain, such as neurons, glial cells, and meninges (1). BT can manifest in individuals across all age groups, spanning from infants to older adults, and possess the potential to exert substantial effects on neurological functioning and overall well-being (2). BT can be categorized as primary or metastatic BT based on its origin. Primary BT originates from cells within the brain, while metastatic BT occurs when cancerous cells from other body parts spread to the brain (3).

Early detection of BT is a crucial factor in enhancing treatment outcomes. Various medical imaging techniques, including computed tomography, single-photon emission computed tomography, positron emission tomography, magnetic resonance spectroscopy, and magnetic resonance imaging (MRI), are employed to gather valuable information regarding the shape, size, location, and metabolic activity of BT, aiding in their diagnosis (4). While these imaging modalities are often combined to yield the most detailed insights into BT, MRI is widely regarded as the gold standard technique due to its excellent soft tissue contrast and widespread availability (3). However, BT exhibits significant heterogeneity, presenting unique variations in size, shape, and location in each individual. Their boundaries are frequently ambiguous or irregular, posing considerable challenges in accurately identifying and segmenting them (3,5).

Precise segmentation of BT plays a pivotal role in enabling healthcare professionals to discern the precise location, dimensions, and morphology of such tumors, while also facilitating an evaluation of their growth patterns and progression (6). Additionally, a precise segmentation mask holds significant potential in aiding surgical planning, postoperative monitoring, and ultimately enhancing the chances of patient survival (7). Furthermore, accurate segmentation of

*Corresponding Author: Tel: +98- 915-359-5578; Email: Hosseini.s.ir@ieee.org

BT assumes particular significance in formulating individualized treatment regimens, encompassing radiation therapy and chemotherapy, thus contributing to improved patient outcomes (6). Many algorithms and techniques have been devised to address the challenge of BT segmentation. These include thresholding methods, region-based segmentation, fuzzy clustering, k-means clustering, and deep learning (DL)-based approaches (8–11). Despite significant advancements in this domain, achieving precise segmentation of BT remains an unresolved matter. In the last decade, deep convolutional neural networks (CNNs) have succeeded remarkably across various domains, particularly in semantic segmentation tasks. Several CNN architectures, including DeepLab (12), RefineNet (13), fully convolutional networks (14), and SegNet, have demonstrated significant advancements in performance and accuracy. These models adopt deep CNN-based image classifiers to the challenging semantic segmentation task.

Accordingly, Ben Naceur et al. (15) aimed to improve BT segmentation in multi-sequence MRI images to facilitate early clinical diagnosis, treatment, and follow-up. They developed three new DL models using end-to-end incremental deep CNNs. These models differed from other CNN-based approaches by incorporating a guided approach to find suitable hyper-parameters instead of relying on trial and error. Moreover, the authors employed ensemble learning to create a more efficient model. Their models were evaluated using the BRATS-2017 dataset, obtaining an average dice similarity coefficient (DSC) of 0.88 over the complete brain region.

Sailunaz et al. (16) developed an automated system with an easy-to-use web app interface for detecting and segmenting BT from MRIs. It has over 90% accuracy and allows healthcare professionals to offer feedback for improved predictions and segmentations.

Havaei et al. (17) utilized deep neural networks for BT segmentation of glioblastomas seen in MRI images, which can have diverse shapes, sizes, and contrasts. They explored CNN architectures that efficiently combine local and global contextual features. They also employed a convolutional implementation of a fully connected layer for faster processing. A 2-phase training procedure was introduced to address tumor label imbalance, and a cascade architecture further enhances performance.

Rehman et al. (18) used a modified SegNet to accurately identify BT in MRI images. Their approach had efficient training and achieved a global accuracy of 99.93%. Mary Cynthia and Merlin Livingston (19) compared four BT segmentation techniques using MRI images to determine the most accurate method. They evaluated Otsu thresholding, the level set method, the fuzzy C-means level set method, and the discrete wavelet transform (DWT) with morphological processing, where they achieved 96% accuracy using the DWT-morphology technique.

Ranjbarzadeh et al. (20) developed a BT segmentation system that utilizes a preprocessing approach to work on smaller image parts, reducing computing time and overfitting. They introduced a cascade to mine local and global features efficiently. They also presented a novel distance-wise attention mechanism to improve segmentation accuracy, where their model achieved mean DSC of 0.9203, 0.9113, and 0.8726 for different tumor regions on the BRATS 2018 dataset. Aggarwal et al. (21) developed an efficient method for BT segmentation using an improved residual network (ResNet). Their approach achieved nearly 85% accuracy on BRATS 2020 MRI data. Aboussaleh et al. (22) utilized a CNN-based approach for simultaneously predicting and segmenting cerebral tumors. They addressed the challenges of specialist intervention, long run-time, and feature extraction using a simple binary annotation for tumor presence. The DL model is trained on the BraTS 2017 dataset with different types of gliomas. Their method achieved 91% accuracy in tumor classification and an 82.35% DSC in tumor segmentation.

Tripathi et al. (23) presented a fully automatic DL method, called CCN-PR-Seg-net, for BT segmentation in MRI images. Their method focused on preserving boundary details of irregular tumor regions, using internal residual connections in the encoder and decoder to avoid information loss. Cross-channel normalization and parametric rectified linear units are employed for a balanced network output. Their model exhibited a global accuracy of 0.998. Gunasekara et al. (24) developed a threefold DL architecture for accurate BT segmentation. It involves a CNN-based classification stage, a region-based CNN for localization stage, and a Chan-Vese segmentation stage. Their architecture achieved an average DSC of 0.92, demonstrating its glioma and meningioma segmentation effectiveness. Haq et al. (25) presented a hybrid approach utilizing deep CNNs and machine learning classifiers to precisely segment and classify BT in MRIs. In the initial stage, a CNN was introduced to learn the feature map of the brain MRI images, with a focus on the tumor marker region. Subsequently, a faster region-based CNN was developed to facilitate tumor region localization, and a region proposal network was employed, achieving a DSC of 97.1% in BT segmentation. Eker et al. (26) conducted a BT segmentation using a dataset with diverse images regarding structural complexity, viewing angles, various device usage, noise, and bias field effects. Importantly, no preprocessing techniques were applied to the dataset. They employed foundational models, specifically U-Net and fully convolutional network, and extended their scope to include transfer learning-based approaches. They incorporated VGG, XceptionNet, InceptionNet, and ResNet architectures as the foundational structures within these models,

wherein they reached a DSC of 0.9169 using the VGG-19.

Sobhaninia et al. (27) proposed a DL-based solution for BT segmentation, which is called LinkNet. They examined diverse angles of MRI, employing various networks for the segmentation process. The efficacy of using distinct networks for MRI image segmentation is assessed through a comparative analysis with results obtained from a single network. Experimental evaluations reveal that a DSC of 0.73 is attained for a single LinkNet model for all directions of MRI images. Badža and Barjaktarović (28) have introduced a compact convolutional neural autoencoder for BT segmentation based on semantic segmentation. Various data division and evaluation methods assessed the model's generalization ability in medical diagnostics. The best results, with a pixel accuracy of 99.28% and a DSC of 72.87%, were achieved using record-wise data division and training with an augmented dataset.

This study's primary contribution lies in applying the ensemble architecture of DeepLabV3+ in conjunction with three established deep CNNs, namely ResNet50, MobileNetV2, and XceptionNet. Should this approach prove successful, it offers significant benefits to partitioners and physicians by enabling precise localization of BT with reduced effort and computational resources compared to existing methods. The rationale for using hybrid methods, such as ensemble DL models, arises from the need to balance accuracy, precision, and robustness in BT segmentation. This task is particularly challenging due to tumor heterogeneity and complex morphology. While some models may report nearly perfect accuracy under ideal conditions, these results do not consistently generalize across diverse cases and datasets. Hybrid methods like the ensemble approach combining MobileNetV2, ResNet50, and XceptionNet within DeepLabV3+ provide multiple perspectives and encoding strengths, reducing model bias and enhancing adaptability across varied axial views in MRI images. This multi-encoder configuration addresses both fine-detail segmentation and generalization, thereby potentially achieving high accuracy levels while minimizing overfitting, making the model better suited for clinical deployment. Such hybrid techniques ensure reliable tumor delineation, which is crucial for improving treatment planning and patient outcomes.

Materials and Methods

Figure 1 depicts the primary framework employed in this study. The MRI dataset was acquired from the Figshare repository and subjected to image processing techniques. Subsequently, an ensemble architecture consisting of three well-known deep CNN models and DeepLabV3+ was utilized for BT segmentation.

Dataset

This study utilized BT's publicly available dataset, consisting of a wide-ranging compilation of 3064 T_1-weighted contrast-enhanced MRI images captured in the form of two–dimensional slices with a resolution of 512×512 pixels in the Nanfang Hospital, Guangzhou, China, and General Hospital, Tianjin Medical University, China, from 2005 to 2010 (29,30).

It was made available online in 2015, and subsequent updates were provided until 2017. This dataset has garnered significant attention in recent MRI tumor analysis research. To ensure the rigor and reliability of our proposed model, we exclusively utilized the most recent version of the dataset, which was last updated in 2017. The provided images were derived from sagittal, coronal, and axial planes and stem from a cohort of 233 patients diagnosed with three specific categories of BT: glioma, meningioma, and pituitary tumor.

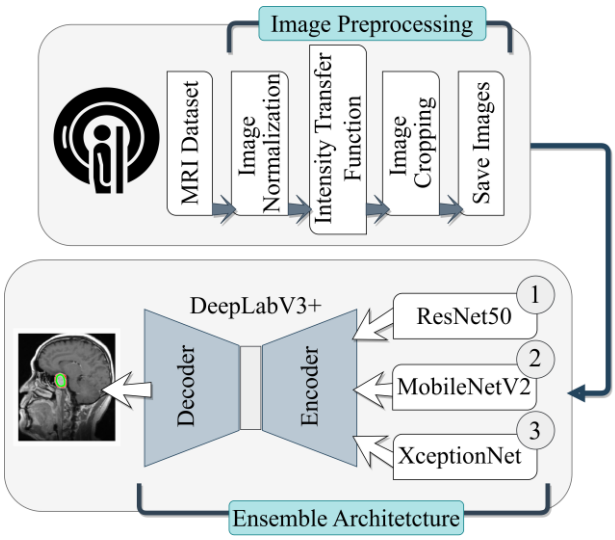


Figure 1. The main framework of this study involves gathering the BT dataset from the Figshare repository and applying image processing techniques. Subsequently, an ensemble architecture comprising CNN and DeepLabV3+ is employed for BT segmentation

The dataset encompasses a total of 1426 slices corresponding to glioma cases, 708 slices originating from individuals with meningioma, and 930 slices representing pituitary tumors. This study considered these three types of tumors (glioma, meningioma, and pituitary) as one class, named tumors, and the brain/other parts as the brain. Subsequently, a binary segmentation was conducted, wherein each image was segmented into the distinct segments of tumors and brain, facilitating a comprehensive dataset analysis.

Image Preprocessing

Before incorporating the images into the deep network training, a few image-processing steps were executed. It is important to note that the dataset is provided in ".mat" format, facilitating access to the data for further analysis and manipulation. Given that the intensity values of the MRI images are contingent upon the specific imaging protocol and scanner utilized, normalization was applied to scale the intensities within the original MRI images to a standardized range between 0 and 255. Subsequently, a linear intensity transformation function (ITF) was employed to enhance the contrast of the images. This function effectively modified the intensity values within the images, resulting in an improved visual differentiation of various image regions based on their intensity levels. ITF is expressed as follows:

$$J = (Scaling\ Factor \times im) + Offset, \quad (1)$$

where im denotes the input image, scaling factor refers to the ratio by which intensity values are multiplied to map the original intensities (Low_{in} to $High_{in}$) to the new range (0–255), which can be expressed as follows:

$$Scaling\ Factor = \frac{High_{out} - Low_{out}}{High_{in} - Low_{in}}. \quad (2)$$

Low_{in} and $High_{in}$ define the designated contrast limits within the input grayscale image that are intended to be associated with specific values in the resultant output image. Low_{in} and $High_{in}$ is the bottom 1% and the top 1% of all pixel values and are calculated as cumulative distribution function (CDF) > 0.01 and $CDF \geq 0.99$, respectively. A CDF is expressed as follows:

$$CDF = \frac{cumsum(N)}{sum(N)}, \quad (3)$$

where N will calculate from the following equation:

$$N = Histogram(^{im}/_{n_{bins}}) \quad (4)$$

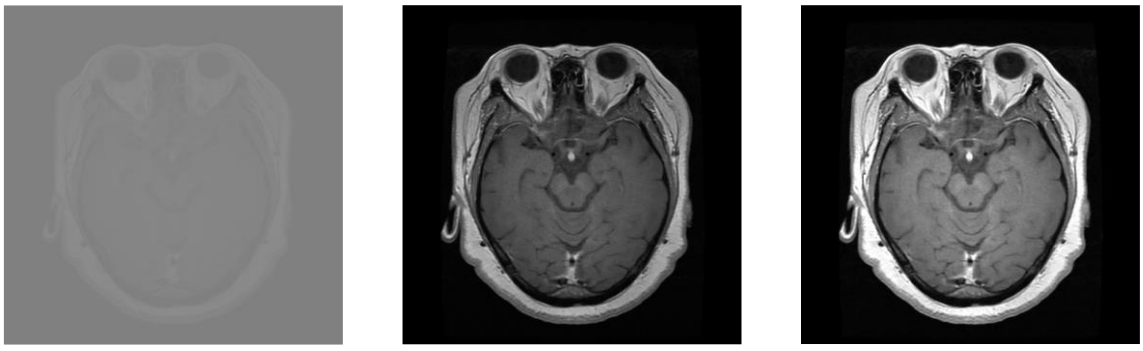
In this equation, $n_{bins} = 256$ if the image is in the form of an 8-bit integer. Correspondingly, Low_{out} and $High_{out}$ represent the contrast limits that are sought for the output grayscale image and are considered as 0 and 255, respectively.

The offset in the linear ITF is used to ensure that the lower bound of the input intensity range (Low_{in}) maps precisely to the desired lower bound of the output range (Low_{out}) after scaling. Offset is expressed as below:

$$\text{Offset} = Low_{out} - (Scaling\ Factor \times Low_{in}) \quad (5)$$

The ITF scales the input image intensity values such that any pixel value below Low_{in} is set to $Low_{out} = 0$, and any pixel value above $High_{in}$ is set to $High_{out} = 255$. This step helps in defining the range of intensities to be adjusted. Figure 2 illustrates the raw, normalized, and adjusted images after applying ITF, showcasing the evident image enhancement achieved through this process.

Figure 3 displays the histogram representation of the images depicted in Figure 2 before and after applying ITF. The histogram serves as a visual depiction of the distribution of pixel intensities within the images. The comparative histograms provide insights into the changes in intensity distribution resulting from ITF, showcasing the alterations in contrast and highlighting the impact of the transformation on the overall image characteristics.



Raw Image Normalized Image Adjusted Image

Figure 2. The sample of raw, normalized, and adjusted image by ITF

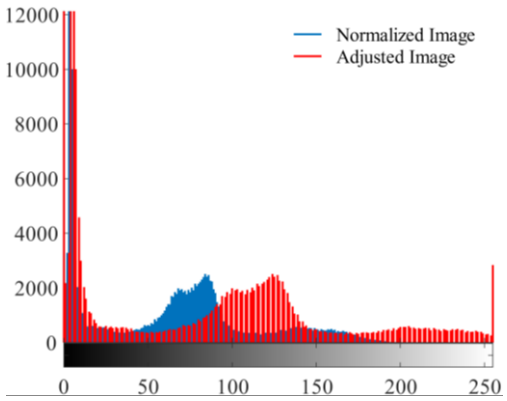


Figure 3. The resultant histogram of the normalized and adjusted image by ITF

In the subsequent stage, an image-cropping process was implemented to confine the region of interest within the boundaries of the brain. Specifically, the extraneous background portions present in the original images were removed by performing a cutout operation. This step ensured that the resulting image exclusively encompassed the relevant brain structures, eliminating background noise or irrelevant image content. The image cropping procedure aimed to enhance the accuracy and specificity of subsequent analyses and interpretations focused solely on the brain region.

Deep Learning

In the past decade, DL applications have surged across diverse healthcare domains, enabling advancements such as neurological disease detection (31,32), breast cancer and automated segmentation of medical images through imaging techniques (33–37), and the identification of neurodevelopmental disorders using non-imaging modalities like electroencephalography (38,39). The ensemble model developed in this study is a pixel-wise segmentation network that leverages the DeepLabV3+ architecture (40), along with three widely recognized pre-trained networks: MobileNetV2 (41), XceptionNet (42), and ResNet50 (43). Combining these components, the ensemble model aims to achieve enhanced performance in pixel-wise segmentation. DeepLabV3+, an extended version of DeepLabV3, is founded on an encoder-decoder framework.

The DeepLabV3+ architecture served as the primary framework for the ensemble model. It is a well-established pixel-wise segmentation network incorporating both an encoder and a decoder. Using convolutional layers, the encoder network effectively captures high-level features and contextual information from the input image. On the other hand, the decoder network refines the feature maps obtained from the encoder through upsampling and skip connections, generating precise pixel-wise segmentation predictions. The DeepLabV3+ architecture was chosen because it handles complex visual patterns and produces accurate segmentations.

Encoder

The encoder component of DeepLabV3+ comprises a deep CNN that commonly employs a modified version of either the widely recognized XceptionNet or MobileNet architecture as its underlying network. This selection of the base network establishes a robust framework for proficient feature extraction, owing to their adeptness in capturing localized and global information. The encoder network incorporates convolutional layers, pooling layers, and activation functions in its architecture. These layers are organized hierarchically, progressively enlarging the network's receptive field and receptive context. As the input image traverses these layers, it undergoes a sequence of convolutions and pooling operations, extracting features across various scales and levels of abstraction.

Several pre-trained networks can be utilized as feature extractors in the encoder component of DeepLabV3+. In our algorithm, we have employed ResNet50, XceptionNet, and MobileNetV2 as potential networks. Specifically, we have adapted and optimized the weights of DeepLabV3+ in combination with these pre-trained CNNs, taking into account the characteristics of our data. The relevant details regarding the image input dimension, number of layers, and number of learnable parameters are briefly presented in Table 1. To accommodate space limitations and avoid confusion, this study does not describe the pre-trained CNN architectures. Their detailed description can be found in their original articles (41-43).

Table 1. The specification of the utilized deep CNN as the encoder part of ensemble architecture

Network	Image input size	Number of layers	Total learnable parameters (millions)
ResNet50	224×224	177	25.5
XceptionNet	299×299	170	22.9
MobileNetV2	224×224	154	3.5

Decoder

The decoder component of the DeepLabV3+ architecture is pivotal in enhancing the encoder's feature maps and generating precise pixel-wise segmentation predictions. Serving as a complementary element to the encoder, it facilitates the model in recovering intricate details and enhancing spatial resolution within the segmentation output. The decoder component of DeepLabV3+ constitutes an integral element aimed at restoring spatial information and generating high-resolution segmentation maps. It achieves this through upsampling and convolutional layers, which enhance the spatial resolution of the feature maps derived from the encoder. These layers progressively upscale the feature maps to match the original dimensions of the input image. The decoder section of DeepLabV3+ encompasses several key components and operations:

Upsampling Layers: These layers employ bilinear interpolation or transposed convolutions to magnify the

feature maps, effectively recovering spatial details and enabling more detailed segmentation predictions.

Skip connections: Establishing connections between corresponding layers in the encoder and decoder sections, skip connections play a vital role in integrating low-level and high-level features. This integration aids in the recovery of fine details and enhances segmentation accuracy. The inclusion of skip connections enables the model to capture both local and global information, contributing to improved segmentation results.

Atrous spatial pyramid pooling (ASPP): The ASPP module, present in both the encoder and decoder parts, connects the two components. It encompasses parallel atrous convolutions with varying dilation rates, allowing for capturing multi-scale contextual information. The output of the ASPP module is typically concatenated with the upsampled feature maps from the decoder, thereby enriching the overall feature representation.

Convolutional layers: Convolutional layers within the decoder component further process and refine the upsampled feature maps. These layers capture intricate details and spatial relationships through a series of convolutional operations, ensuring the segmentation predictions are accurate and precise.

Final segmentation output: The output of the decoder part is a high-resolution segmentation map. Each pixel in this map corresponds to a predicted class or label, providing a pixel-wise segmentation prediction. In applications such as BT segmentation, this output delineates the regions of interest, such as the tumor region.

By incorporating the decoder part in DeepLabV3+, the model effectively recovers spatial details and generates fine-grained segmentation predictions by leveraging the hierarchical information captured by the encoder. The combination of upsampling layers, skip connections, ASPP, and convolutional layers contribute to refining the feature maps, enhancing spatial resolution, and improving the overall accuracy of the segmentation output.

Training Options

The computations and implementation of the proposed ensemble model were carried out using MATLAB R2022b software on a machine equipped with an Intel Core i7 11700k CPU, 32GB of DDR4 memory, MATLAB R2022, and Nvidia GEFORCE RTX 3070 graphics cards. The selection of hyperparameters in this study was accomplished through a systematic trial-and-error procedure to optimize these hyperparameters, where the optimal hyperparameters were determined: stochastic gradient descent with momentum using learning rate of 0.001, a momentum of 0.9, L2 regularization of 0.0005, a maximum epoch of 2, a mini-batch size of 32, and CrossEntropy as a loss function. In addition, an early-stopping criterion was incorporated into our training process. The training was halted if the validation accuracy did not improve over three consecutive validation steps. This approach helps prevent overfitting and ensures the model is trained until it reaches its optimal performance.

The dataset was partitioned into three subsets for specific purposes: 70% for training, 20% for testing, and 10% for validation. This division allows for effective model training, unbiased evaluation of unseen data during testing, and validation to fine-tune hyperparameters and monitor the model's performance. Various metrics, including global accuracy, mean accuracy, intersection over union (IoU), weighted IoU, boundary F1 score (BF), and DSC, were used to evaluate the model's segmentation performance.

Results

To assess the performance of the proposed model, a subset of 20% of the entire dataset was designated as the test data. Several metrics were then calculated to evaluate the model's global performance in segmentation. These metrics include global and mean accuracy, average and weighted IoU, BF, and DSC. BF is a metric that measures the alignment between the predicted boundary of each class and the actual boundary. It indicates how well the predicted boundary corresponds to the actual boundary and is calculated through the following equation:

$$BF = 2 * Precision * Recall / Recall + Precision \quad (6)$$

In Equation (4), precision is computed by dividing the number of true positives by the sum of true positives and false positives. This calculation measures the proportion of correctly identified positive instances out of all the instances predicted as positive. Recall, also known as sensitivity, is calculated by dividing the number of true positives by the sum of true positives and false negatives. This calculation quantifies the proportion of actual positive instances that the model correctly identifies out of all the positive instances present in the data.

Global accuracy is a metric that calculates the ratio of correctly classified pixels, irrespective of class, to the total number of pixels. It provides a quick and computationally efficient estimation of the percentage of pixels that are classified correctly. IoU also referred to as the Jaccard similarity coefficient (JSC), was utilized to assess the percentage of similarity between the predicted labels generated by the model and the ground truth (GT) labels for each of the two classes under consideration. IoU is expressed as follows:

$$IoU, JSC = (Pred \cap GT)/(Pred \cup GT) \quad (7)$$

where $Pred$ is the predicted labels by model. DSC is then calculated using the following equation:

$$DSC = 2 * JSC/1 + JSC \quad (8)$$

The average IoU of each class, weighted by the number of pixels in that class, is a metric that takes into account the disproportionate sizes of classes in an image. It reduces the impact of errors in smaller classes on the overall quality score. This metric provides a more balanced assessment of the IoU performance by assigning higher weightage to

classes with larger pixels.Table 2 presents the calculated metrics as both the average and standard deviation values across the entire test set. The average values provide an overall measure of the performance metrics, while the standard deviation reflects the degree of variation or dispersion among the calculated metrics.

The evaluation metrics, reported as mean ± standard deviation across the entire test set, provide insights into the performance of each model. The global accuracy measures the overall accuracy of the models in correctly classifying tumor and brain regions. Among the models, MobileNetV2 demonstrated the highest global accuracy with a score of 0.997 ± 0.001. ResNet50 closely followed with a global accuracy of 0.997 ± 0.002, while XceptionNet achieved a slightly lower value of 0.995 ± 0.004. The mean accuracy metric indicates the average accuracy of the models in segmenting different tumor regions. MobileNetV2 exhibited the highest mean accuracy of 0.945 ± 0.044, followed by ResNet50 with a value of 0.886 ± 0.072. XceptionNet yielded a mean accuracy of 0.837 ± 0.164. The mean IoU metric quantifies the overlap between the predicted and GT tumor regions. MobileNetV2 achieved a mean IoU of 0.885 ± 0.037, while ResNet50 and XceptionNet obtained scores of 0.859 ± 0.069 and 0.786 ± 0.143, respectively.

The weighted IoU was calculated by weighting the IoU scores based on class frequencies to account for class imbalance. MobileNetV2 attained the highest weighted IoU of 0.995 ± 0.003, followed by ResNet50 with a score of 0.995 ± 0.004. XceptionNet achieved a slightly lower weighted IoU of 0.991 ± 0.008. Lastly, the mean BF metric measures the accuracy of the models in segmenting tumor boundaries. MobileNetV2 exhibited the highest mean BF of 0.931 ± 0.074, followed by ResNet50 with a value of 0.908 ± 0.087. XceptionNet yielded a mean BF of 0.864 ± 0.146. The results indicate that MobileNetV2 outperformed the other models in terms of global accuracy, mean accuracy, mean IoU, weighted IoU, and mean BF. ResNet50 and XceptionNet also performed well, although with slightly lower scores in most metrics compared to MobileNetV2. These findings suggest that employing MobileNetV2 as a backbone of DeeolabV3+ is a promising model for accurate BT segmentation.

Table 2. The result of BT segmentation for the test set. Results are presented as average ± standard deviation

Model	Global accuracy	Mean accuracy	Mean IoU	Weighted IoU	Mean BF	DSC
MobileNetV2	0.997±0.001	0.945±0.044	0.885±0.037	0.995±0.003	0.931±0.074	0.938±0.021
ResNet50	0.997±0.002	0.886±0.072	0.859±0.069	0.995±0.004	0.908±0.087	0.922±0.043
XceptionNet	0.995±0.004	0.837±0.164	0.786±0.143	0.991±0.008	0.864±0.146	0.872±0.097

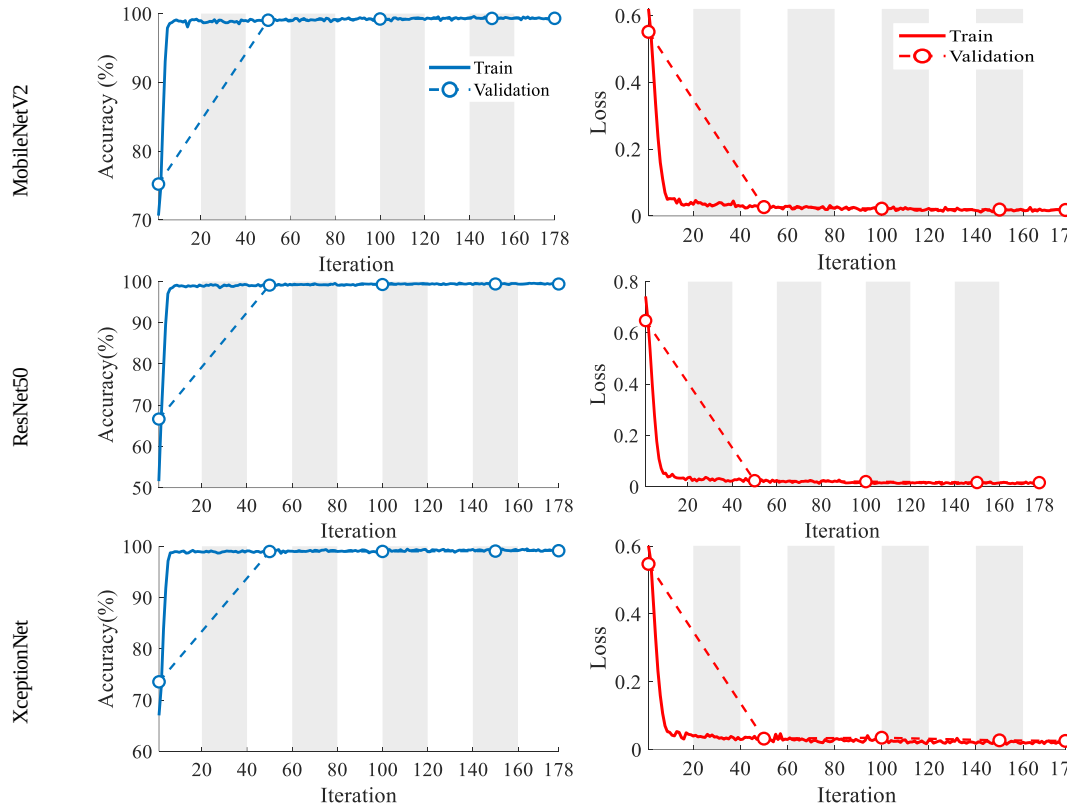


Figure 4. The global accuracy and loss function of the three ensemble models utilized in the study at various iterations, where the loss function is considered as binary cross entropy

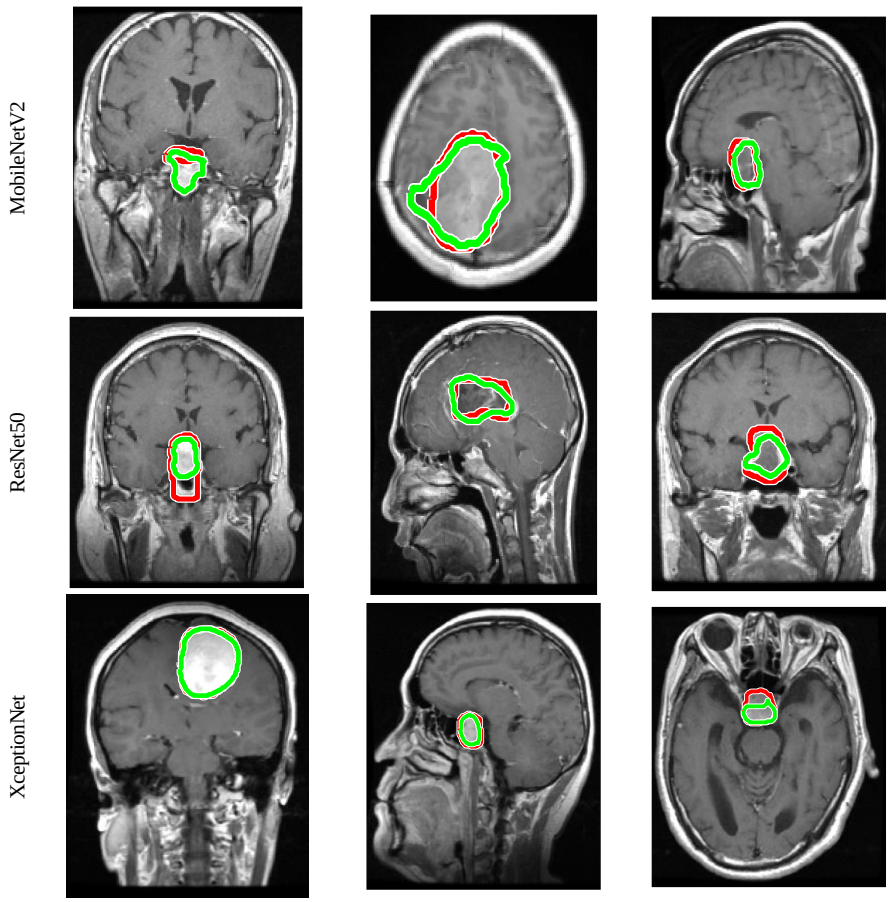


Figure 5. The resultant BT-segmented images from different axial views. The red contour represents GT, while the green contour illustrates the pixel-wise segmented region predicted by the ensemble models.

Figure 4 depicts the global accuracy and loss function of the three ensemble models utilized in the study at various iterations, where the loss function is considered as binary cross entropy.The global accuracy in all three models shows an upward trend with increasing iterations until it reaches a plateau, indicating that further iterations do not lead to significant improvements in accuracy. Despite implementing an early-stopping criterion, all models were trained for a total of 2 epochs or 178 iterations. Additionally, the loss function in all three models demonstrates a consistent downward trend as the iterations progress. This downward trend indicates that the models' training process is converging towards the potential minimum of the loss function. As the iterations continue, the loss gradually decreases until it reaches a point where further iterations have a negligible impact on reducing the loss. This suggests that the models have achieved a relatively optimal state in terms of minimizing the loss function.

Figure 5 illustrates a sample of BT-segmented images from different axial views, showcasing the capabilities of the ensemble models employed for BT segmentation. In this depiction, the GT is represented by the red contour, while the green contour illustrates the pixel-wise segmented region predicted by the ensemble models. The visual representation exemplifies the potential and effectiveness of the utilized ensemble models in accurately delineating BT regions.

Discussion

This research focuses on developing and optimizing an ensemble comprising DL-based algorithms to segment BT regions in MRI images. The primary objective of this study is to achieve accurate and reliable segmentation results for tumor regions across a range of images obtained from different axial planes. The adopted DeepLabV3+ architecture leverages the ASPP convolutional structure, offering distinct advantages in generating dense feature maps and preserving spatial information. This property improves localization accuracy, consequently impacting the overall segmentation process. The favorable outcomes reported in previous studies and the present work reaffirm the efficacy of the ASPP module in achieving successful segmentation outcomes for various tasks (44,45).

By employing the ASPP module, the DeepLabV3+ architecture enhances the feature extraction process by incorporating multiple parallel atrous convolutions at different dilation rates. This enables the model to capture multi-scale contextual information effectively, allowing for a more comprehensive understanding of the image and its semantic content. The resulting dense feature maps obtained from the ASPP module contribute

to the precise localization of object boundaries and boundaries of interest, which is crucial for accurate segmentation. The combination of MobileNetV2 as the encoder and DeepLabV3+ as the segmentation framework has yielded impressive results, with the highest achieved accuracy among the evaluated models (44). MobileNetV2 is a lightweight and efficient CNN architecture. It is specifically designed for mobile and embedded devices to balance model size and computational efficiency. The MobileNetV2 architecture incorporates depthwise separable convolutions and linear bottlenecks to reduce the number of parameters while maintaining expressive power. These design choices result in a compact network that can efficiently process input images (46).

When employed as the encoder within the DeepLabV3+ model, MobileNetV2 exhibits exceptional proficiency in extracting high-level features and capturing contextual information from BT images. The utilization of depth-wise separable convolutions facilitates efficient feature extraction, while the integration of linear bottlenecks aids in preserving crucial spatial information. This amalgamation plays a pivotal role in ensuring precise localization and segmentation of tumors, contributing to the model's overall accuracy.

The findings derived from our experimental analysis unequivocally demonstrate the superiority of the MobileNetV2-DeepLabV3+ combination over the alternative models considered in this study. The obtained accuracy surpasses that of competing architectures, including ResNet50 and XceptionNet. This outcome serves as a testament to the efficacy of MobileNetV2 in effectively addressing the intricacies associated with BT segmentation. Moreover, integrating the DeepLabV3+ framework further enhances the process by enabling meticulous refinement and precise delineation of tumor boundaries. Consequently, this combination proves to be highly effective in achieving accurate and reliable BT segmentation results, which is of great significance in clinical applications.

The exceptional performance of the MobileNetV2-DeepLabV3+ combination can be attributed to the synergistic integration of their respective strengths. MobileNetV2 excels in efficient feature extraction, effectively extracting relevant and discriminative features from BT images. On the other hand, DeepLabV3+ demonstrates remarkable capabilities in semantic segmentation, enabling the model to accurately classify and delineate tumor regions based on the extracted features.

By leveraging the efficient feature extraction capabilities of MobileNetV2 and the semantic segmentation prowess of DeepLabV3+, our model exhibits a comprehensive understanding of the underlying tumor structures. This combination enables the model to capture fine-grained details of the tumors, facilitating accurate segmentation and localization. The efficient feature extraction ensures that relevant information is preserved, while the semantic segmentation techniques refine the segmentation boundaries with high precision.

The synergy between MobileNetV2 and DeepLabV3+ allows our model to leverage the strengths of both architectures, leading to the effective capture of intricate tumor characteristics and the achievement of accurate segmentation outcomes. This comprehensive understanding of tumor structures is crucial for accurate diagnosis and treatment planning in BT analysis. Table 3 provides a concise overview of a comparative analysis conducted in conjunction with several state-of-the-art studies that employed the same dataset. Notably, our model demonstrated superior performance compared to the alternative models assessed in these studies. Our result demonstrates an approximate 20% superiority over the outcomes reported in references (47–49) concerning DSC. Furthermore, our findings exhibit competitive performance closely aligned with the results reported in references (50) and (51). A higher DSC signifies more significant overlap and agreement between the predicted and GT segmentation masks. This indicates that our model's segmentation results align more closely with the manually annotated GT data, reflecting improved accuracy and precision in identifying tumor regions within the brain images. This higher DSC suggests enhanced performance and reliability in our segmentation approach compared to the studies cited in the references (47–49). In terms of accuracy, our results demonstrate competitiveness with the accuracy reported in references (49–51), underscoring the efficacy of DL methods in the challenging task of BT segmentation. This suggests that our model's accuracy is comparable to and potentially higher than that achieved by the mentioned models.

Table 3. Comparative analysis of our performance with other state-of-art studies that utilized the same dataset

Ref	Year	Model	Performance
(27)	2018	DL-based solution, named LinkNet	DSC of 0.730
(18)	2019	Deep autoencoder-decoder framework for semantic segmentation, named SegNet	DSC of 0.931 Accuracy of 0.934
(28)	2021	compact convolutional neural autoencoder	DSC of 0.728 Accuracy of 0.992
(24)	2021	DL and active contouring	DSC of 0.920 Accuracy of 0.945
(26)	2023	VGG-19 and fully connected network	DSC of 0.916
Proposed model		Ensemble architecture of DeepLabV3+ and MobileNetV2	DSC of 0.938 Accuracy of 0.997

Despite the excellent result achieved, it is pertinent to acknowledge certain limitations that warrant consideration. Firstly, while exhibiting commendable accuracy, the proposed model might face challenges in scenarios with highly heterogeneous or rare tumor subtypes not extensively represented in the current dataset. This indicates the need for further investigation on a more diverse dataset encompassing a broader spectrum of BT variations. Moreover, despite the robustness demonstrated by our method, there exists room for improvement in terms of interpretability and explainability. Integrating interpretability features into the model architecture would contribute to a more transparent decision-making process, facilitating better clinician understanding and trust.

Conclusion

To conclude, the study's primary contribution lies in addressing the challenges of BT segmentation by leveraging the strengths of efficient feature extraction and semantic segmentation techniques. This study presents a robust ensemble architecture integrating DeepLabV3+ with three prominent deep CNNs: MobileNetV2, ResNet50, and XceptionNet, to segment BT in MRI images accurately. The proposed models are validated using the Figshare BT dataset, wherein the MobileNetV2-DeepLabV3+ combination demonstrated superior performance, achieving an accuracy of 0.997 and DSC of 0.938, showcasing its efficacy in capturing intricate tumor characteristics and achieving accurate segmentation outcomes. It also offers significant benefits to partitioners and physicians by enabling precise localization of BT with reduced effort and computational resources compared to existing methods.

The findings of this study contribute to the ongoing efforts to advance the field of BT segmentation, with implications for improved clinical diagnosis and treatment planning. As medical imaging technologies and DL methodologies evolve, the proposed ensemble architecture is a promising step towards more accurate and efficient BT segmentation, aiding healthcare professionals in their critical decision-making processes.

Future endeavors could explore integrating vision transformer models and DL architectures enriched with attention mechanisms to address these limitations and advance the field. Vision transformers have shown promise in capturing long-range dependencies in images, potentially benefiting the segmentation of intricate and diverse BT patterns. Also, incorporating attention blocks into the model could further refine the focus on relevant regions, enhancing the model's interpretability and adaptability. Furthermore, given the inherent challenge of class imbalance in medical image datasets, future work should consider implementing advanced techniques such as ensemble learning or focal loss to address this issue, thereby enhancing the model's efficacy across diverse tumor types.

Acknowledgment

This study was supported in part by grant No 17171 from the National Ethical Committee in Biomedical Research of Health and by a teaching and research scholarship from the Shiraz University of Medical Sciences (Dr Fariba Zarei).

References

1. Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, et al. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. Acta Neuropathol. 2016 Jun 9;131(6):803–20.
2. Arora RS, Alston RD, Eden TOB, Estlin EJ, Moran A, Birch JM. Age–incidence patterns of primary CNS tumors in children, adolescents, and adults in England. Neuro Oncol. 2009 Aug 1;11(4):403–13. doi: 10.1215/15228517-2008-097
3. Işın A, Direkoğlu C, Şah M. Review of MRI-based Brain Tumor Image Segmentation Using Deep Learning Methods. Procedia Comput Sci. 2016;102:317–24. doi: 10.1016/j.procs.2016.09.407
4. Modaresnia Y, Abedinzadeh Torghabeh F, Hosseini SA. EfficientNetB0′s Hybrid Approach for Brain Tumor Classification from MRI Images Using Deep Learning and Bagging Trees. In: 2023 13th International Conference on Computer and Knowledge Engineering, ICCKE 2023. IEEE; 2023. p. 234–9. doi:10.1109/ICCKE60553.2023.10326290
5. Zhang W, Wu Y, Yang B, Hu S, Wu L, Dhelim S. Overview of Multi-Modal Brain Tumor MR Image Segmentation. Healthcare. 2021 Aug 16;9(8):1051. doi: 10.3390/healthcare9081051
6. Fawzi A, Achuthan A, Belaton B. Brain Image Segmentation in Recent Years: A Narrative Review. Brain Sci. 2021 Aug 10;11(8):1055. doi: 10.3390/brainsci11081055
7. Liu Z, Tong L, Chen L, Jiang Z, Zhou F, Zhang Q, et al. Deep learning based brain tumor segmentation: a survey. Complex & Intelligent Systems. 2023 Feb 9;9(1):1001–26.
8. Kumar A. Study and analysis of different segmentation methods for brain tumor MRI application. Multimed Tools Appl. 2023 Feb 16;82(5):7117–39.
9. Angulakshmi M, Lakshmi Priya GG. Automated brain tumour segmentation techniques- A review. Int J Imaging Syst Technol. 2017 Mar;27(1):66–77.
10. Aboussaleh I, Riffi J, Mahraz AM, Tairi H. Brain Tumor Segmentation Based on Deep Learning′s Feature Representation. J Imaging. 2021 Dec 8;7(12):269.
11. Ullah F, Salam A, Abrar M, Amin F. Brain Tumor Segmentation Using a Patch-Based Convolutional Neural Network: A Big Data Analysis Approach. Mathematics. 2023 Mar 28;11(7):1635.
12. Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Trans Pattern Anal Mach Intell. 2018 Apr 1;40(4):834–48.
13. Lin G, Milan A, Shen C, Reid I. RefineNet: Multi-path Refinement Networks for High-Resolution

Semantic Segmentation. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE; 2017. p. 5168–77.
14. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE; 2015. p. 3431–40.
15. naceur M Ben, Saouli R, Akil M, Kachouri R. Fully Automatic Brain Tumor Segmentation using End-To-End Incremental Deep Neural Networks in MRI images. Comput Methods Programs Biomed. 2018 Nov;166:39–49.
16. Sailunaz K, Bestepe D, Alhajj S, Özyer T, Rokne J, Alhajj R. Brain tumor detection and segmentation: Interactive framework with a visual interface and feedback facility for dynamically improved accuracy and trust. Raja G, editor. PLoS One. 2023 Apr 17;18(4):e0284418.
17. Havaei M, Davy A, Warde-Farley D, Biard A, Courville A, Bengio Y, et al. Brain tumor segmentation with Deep Neural Networks. Med Image Anal. 2017 Jan;35:18–31.
18. Rehman A, Naz S, Naseem U, Razzak I, Hameed IA. Deep AutoEncoder-Decoder Framework for Semantic Segmentation of Brain Tumor. Aust J Intell Inf Process Syst. 2019;15:53–60.
19. Mary Cynthia S, Merlin Livingston LM. Brain Tumour Segmentation Methods Based on DWT. In 2022. p. 489–96.
20. Ranjbarzadeh R, Bagherian Kasgari A, Jafarzadeh Ghoushchi S, Anari S, Naseri M, Bendechache M. Brain tumor segmentation based on deep learning and an attention mechanism using MRI multi-modalities brain images. Sci Rep. 2021 May 25;11(1):10930.
21. Aggarwal M, Tiwari AK, Sarathi MP, Bijalwan A. An early detection and segmentation of Brain Tumor using Deep Neural Network. BMC Med Inform Decis Mak. 2023 Apr 26;23(1):78.
22. Aboussaleh I, Riffi J, Mahraz AM, Tairi H. Brain Tumor Segmentation Based on Deep Learning′s Feature Representation. J Imaging. 2021 Dec 8;7(12):269.
23. Tripathi S, Verma A, Sharma N. Automatic segmentation of brain tumour in MR images using an enhanced deep learning approach. Comput Methods Biomech Biomed Eng Imaging Vis. 2021 Mar 4;9(2):121–30.
24. Gunasekara SR, Kaldera HNTK, Dissanayake MB. A Systematic Approach for MRI Brain Tumor Localization and Segmentation Using Deep Learning and Active Contouring. Yao Y h., editor. J Healthc Eng. 2021 Feb 28;2021:1–13.
25. Haq EU, Jianjun H, Huarong X, Li K, Weng L. A Hybrid Approach Based on Deep CNN and Machine Learning Classifiers for the Tumor Segmentation and Classification in Brain MRI. Koundal D, editor. Comput Math Methods Med. 2022 Aug 8;2022:1–18.
26. Eker AG, Pehlivanoğlu MK, İnce İ, Duru N. Deep Learning and Transfer Learning Based Brain Tumor Segmentation. In: 2023 8th International Conference on Computer Science and Engineering (UBMK). IEEE; 2023. p. 163–8.
27. Sobhaninia Z, Rezaei S, Noroozi A, Ahmadi M, Zarrabi H, Karimi N, et al. Brain Tumor Segmentation Using Deep Learning by Type Specific Sorting of Images. 2018.
28. Badža MM, Barjaktarović MČ. Segmentation of Brain Tumors from MRI Images Using Convolutional Autoencoder. Applied Sciences. 2021 May 10;11(9):4317.
29. Cheng J, Huang W, Cao S, Yang R, Yang W, Yun Z, et al. Enhanced Performance of Brain Tumor Classification via Tumor Region Augmentation and Partition. Zhang D, editor. PLoS One. 2015 Oct 8;10(10):e0140381.
30. Cheng J, Yang W, Huang M, Huang W, Jiang J, Zhou Y, et al. Retrieval of Brain Tumors by Adaptive Spatial Pooling and Fisher Vector Representation. Yap PT, editor. PLoS One. 2016 Jun 6;11(6):e0157112.
31. Abedinzadeh Torghabeh F, Hosseini SA, Ahmadi Moghadam E. Enhancing Parkinson's disease severity assessment through voice-based wavelet scattering, optimized model selection, and weighted majority voting. Med. Nov. Technol. Devices. 2023 Dec 1;20:100266. doi: 10.1016/j.medntd.2023.100266
32. Abedinzadeh Torghabeh F, Ahmadi Moghadam E, Hosseini SA. Simultaneous time-frequency analysis of gait signals of both legs in classifying neurodegenerative diseases. Gait Posture. 2024 Sep;113:443–51. doi: 10.1016/j.gaitpost.2024.07.302
33. Rezaei S, Zadeh HG, Gholizadeh MH, Fayazi A, Rezaee K. Modeling and Predicting the Survival of Breast Cancer Patients via Deep Neural Networks and Bayesian Algorithm. Iranian Journal of Medical Physics. 2024;21(3):203–10. doi: 10.22038/ijmp.2023.69096.2217
34. Abedinzadeh Torghabeh F, Modaresnia Y, Hosseini SA. An Efficient Approach for Breast Abnormality Detection through High-Level Features of Thermography Images. In: 2023 13th International Conference on Computer and Knowledge Engineering, ICCKE 2023. IEEE; 2023. p. 54–9. doi: 10.1109/ICCKE60553.2023.10326246
35. Mirzaei F, Parishan M, Faridafshin M, Faghihi R, Sina S. Automated Tumor Segmentation Based on Hidden Markov Classifier using Singular Value Decomposition Feature Extraction in Brain MR images. Iranian Journal of Medical Physics. 2018;15(Special Issue-12th. Iranian Congress of Medical Physics):184. doi: 10.22038/ijmp.2018.12797
36. Firouzmand M, Majidzadeh K, Jafari M, Haghighat S, Esmaeili R, Moradi L, et al. A Framework for Promoting Passive Breast Cancer Monitoring: Deep Learning as an Interpretation Tool for Breast Thermograms. Iranian Journal of Medical Physics. 2024;21(4):237–48. doi: 10.22038/ijmp.2023.71683.2268
37. Modaresnia Y, Abedinzadeh Torghabeh F, Hosseini SA. Enhancing multi-class diabetic retinopathy detection using tuned hyper-parameters and modified deep transfer learning. Multimed Tools

Appl. 2024 Mar 8:1-22. doi: 10.1007/s11042-024-18506-3
38. Ahmadi Moghadam E, Abedinzadeh Torghabeh F, Hosseini SA, Moattar MH. Improved ADHD Diagnosis Using EEG Connectivity and Deep Learning through Combining Pearson Correlation Coefficient and Phase-Locking Value. Neuroinformatics. 2024 Oct 18. doi: 10.1007/s12021-024-09685-3
39. Abedinzadeh Torghabeh F, Modaresnia Y, Hosseini SA. EEG-Based Effective Connectivity Analysis for Attention Deficit Hyperactivity Disorder Detection Using Color-Coded Granger-Causality Images and Custom Convolutional Neural Network. International Clinical Neuroscience Journal. 2023;10(November):e12. doi: 10.34172/icnj.2023.12
40. Chen LC, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. 2018.
41. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In: CVPR. 2018.
42. Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions. 2017.
43. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2015.
44. Ding P, Qian H, Zhou Y, Yan S, Feng S, Yu S. Real-time efficient semantic segmentation network based on improved ASPP and parallel fusion module in complex scenes. J Real Time Image Process. 2023 Jun 6;20(3):41.
45. Lian X, Pang Y, Han J, Pan J. Cascaded hierarchical atrous spatial pyramid pooling module for semantic segmentation. Pattern Recognit. 2021 Feb;110:107622.
46. Memon MM, Hashmani MA, Junejo AZ, Rizvi SS, Raza K. Unified DeepLabV3+ for Semi-Dark Image Semantic Segmentation. Sensors. 2022 Jul 15;22(14):5312.
47. Eker AG, Pehlivanoğlu MK, İnce İ, Duru N. Deep Learning and Transfer Learning Based Brain Tumor Segmentation. In: 2023 8th International Conference on Computer Science and Engineering (UBMK). IEEE; 2023. p. 163–8. doi: 10.1109/UBMK59864.2023.10286591
48. Sobhaninia Z, Rezaei S, Noroozi A, Ahmadi M, Zarrabi H, Karimi N, et al. Brain Tumor Segmentation Using Deep Learning by Type Specific Sorting of Images [Internet]. 2018. Available from: http://arxiv.org/abs/1809.07786
49. Badža MM, Barjaktarović MČ. Segmentation of Brain Tumors from MRI Images Using Convolutional Autoencoder. Applied Sciences. 2021 May 10;11(9):4317. doi: 10.3390/app11094317
50. Rehman A, Naz S, Naseem U, Razzak I, Hameed IA. Deep AutoEncoder-Decoder Framework for Semantic Segmentation of Brain Tumor. Australian Journal of Intelligent Information Processing Systems. 2019;15(4):54–60.
51. Gunasekara SR, Kaldera HNTK, Dissanayake MB. A Systematic Approach for MRI Brain Tumor Localization and Segmentation Using Deep Learning and Active Contouring. Yao Y h., editor. J Healthc Eng. 2021 Feb 28;2021:1–13. doi: 10.1155/2021/6695108