Konya Mühendislik Bilimleri Dergisi, c.14, s.3, 1648-1666, 2026
Konya Journal of Engineering Sciences, v.14, n.3, 1648-1666, 2026
ISSN: 2667-8055 (Electronic)
DOI: 10.36306/konjes.1747393

BRAIN TUMOR SEGMENTATION USING U-NET-BASED DEEP LEARNING MODELS

[1]Merve KAYAR, [2,]*Emine OKUMUŞ, [3]Rıdvan SARAÇOĞLU

[1]*Van Yuzuncu Yil University, Institute of Natural and Applied Sciences, Artificial Intelligence and Robotics, Van, Türkiye*
[2]*Van Yuzuncu Yil University, Faculty of Engineering, Department of Food Engineering, Van, Türkiye*
[3]*Van Yuzuncu Yil University, Faculty of Engineering, Department of Computer Engineering, Van, Türkiye*
[1]mrvekyr@gmail.com, [2]emineokumus@yyu.edu.tr, [3]ridvansaracoglu@yyu.edu.tr

Highlights

- Four CNN-based U-Net architectures were developed for brain tumor segmentation
- Models were trained and tested using BraTS 2019 and Figshare datasets
- U-Net++ achieved the best performance with the highest Dice score and Specificity
- Classic U-Net, U-Net+ResNet50, and U-Net+VGG-16 followed in segmentation accuracy
- Using 3D U-Net variants and data augmentation is recommended for future studies

***Corresponding Author:** Emine OKUMUŞ, emineokumus@yyu.edu.tr

Konya Mühendislik Bilimleri Dergisi, c.14, s.3, 1648-1666, 2026
Konya Journal of Engineering Sciences, v.14, n.3, 1648-1666, 2026
ISSN: 2667-8055 (Electronic)
DOI: 10.36306/konjes.1747393

BRAIN TUMOR SEGMENTATION USING U-NET-BASED DEEP LEARNING MODELS

[1]Merve KAYAR, [2,]*Emine OKUMUŞ, [3]Rıdvan SARAÇOĞLU

[1]Van Yuzuncu Yil University, Institute of Natural and Applied Sciences, Artificial Intelligence and Robotics, Van, Türkiye
[2]Van Yuzuncu Yil University, Faculty of Engineering, Department of Food Engineering, Van, Türkiye
[3]Van Yuzuncu Yil University, Faculty of Engineering, Department of Computer Engineering, Van, Türkiye
[1]mrvekyr@gmail.com, [2]emineokumus@yyu.edu.tr, [3]ridvansaracoglu@yyu.edu.tr

(Received: 21.07.2025; Accepted in Revised Form: 18.02.2026)

ABSTRACT: When examining segmentation methods developed for brain tumor detection, a significant gap exists in the comparative performance analysis of standard and hybrid U-Net models. This study systematically evaluates four Convolutional Neural Network (CNN) models to address this gap, particularly regarding the practical effectiveness of integrating complex, pre-trained encoders: Classic U-Net, U-Net+VGG-16, U-Net+ResNet50, and U-Net++. The BraTS 2019 dataset was used for training and testing the models (with 5-fold cross-validation), and the Figshare dataset was used for additional validation, performing tumor segmentation and generating predictions. U-Net++ outperformed both the conventional U-Net and the hybrid models, indicating its higher segmentation performance under the conditions evaluated in this study. While U-Net++ achieved superior results in Dice coefficient, sensitivity, specificity, and Jaccard Index values (0.920, 0.890, 0.973, and 0.850, respectively), the U-Net+VGG-16 hybrid model (0.697, 0.659, 0.862, and 0.535, respectively) and the U-Net+ResNet50 hybrid model (0.585, 0.550, 0.800, and 0.414) produced lower Dice scores. When we examine the primary reasons for this poor performance output, we identify and discuss these as the incompatibility of 2D encoder blocks with 3D medical data and the lack of unoptimized transfer learning. We believe this finding is of critical importance for future model designs. Our study demonstrates that the U-Net++ architecture, with its dense and intertwined skip connections, provides a more effective solution than simple hybrid models built with powerful classifiers such as VGG-16 and ResNet50. This allows us to obtain extremely important results for precise brain tumor segmentation.

Keywords: *Brain Tumor, Computer-Aided Diagnosis, Convolutional Neural Networks, Image Processing, Magnetic Resonance, U-Net.*

1. INTRODUCTION

Brain tumors are one of the most serious health problems that must be diagnosed early for effective treatment [1]. Magnetic Resonance Imaging (MRI) serves as a fundamental tool in brain tumor imaging, but manual detection is time-consuming and variable [2]. Therefore, the development of deep learning-enabled automated computer diagnosis (CAD) systems has accelerated.

Image segmentation is a crucial first step in automated computer diagnosis systems [3]. The transition from traditional machine learning, which relies on hand-crafted features, to deep learning, particularly Fully Convolutional Networks (FCNs), has enabled end-to-end pixel-by-pixel classification, significantly improving segmentation accuracy [4, 5]. When examining the U-Net architecture [6], it is seen that its symmetric encoder-decoder structure and skip connections, combined with precise localization and efficient incorporation of contextual information, make it an important structure for biomedical image segmentation.

***Corresponding Author:** Emine OKUMUŞ, emineokumus@yyu.edu.tr

The success of U-Net has also facilitated numerous new studies. 3D U-Net [7] extended the U-Net architecture to volumetric data to capture 3D contextual information. Other work to improve feature extraction explores hybrid models that integrate pre-trained networks such as VGG-16 [8] and ResNet [9] as robust encoders [10, 11]. Another study introduced U-Net++ [12], which introduces densely intertwined skip connections to reduce the semantic gap between encoder and decoder feature maps through architectural improvements, promising more accurate segmentation.

Previous studies have predominantly focused on computationally intensive 3D models, whereas the potential limitations of hybrid approaches across different application contexts have received comparatively limited attention. However, a direct comparison of hybrid variants of 2D U-Net and advanced architectures such as U-Net++ for tumor segmentation is still needed.

This study aims to provide this critical comparison. In this study, we examine four models: Classic U-Net, U-Net+VGG-16, U-Net+ResNet50, and U-Net++, to identify the most effective architecture and analyze the reasons behind their performance trends. Our study contributes to: 1) benchmarking these models separately on the BraTS 2019 and Figshare datasets, 2) analyzing why U-Net++ outperforms hybrid models, and 3) providing insights that will guide model design for future medical image analysis.

2. MATERIAL AND METHODS

BraTS 2019 dataset [13] was used in both the training and evaluation phases of this study (Project No. 2021/05-28 by Van Yuzuncu Yil University (Turkey) Non-Interventional Clinical Research Ethics Committee). BraTS 2019 is a publicly available dataset and has undergone extensive preprocessing before being made available [13]. This standard preprocessed format provided by the BraTS 2019 dataset was used directly; no additional preprocessing was performed by the study authors. The dataset contains 285 brain images, 210 of which are HGG and 75 of which are LGG. Each patient has four different MRI scans: T1, T2, T2 Flair, and T1-enhanced (T1c). These images are three-dimensional matrices, each measuring 240x240x155. All images underwent significant pre-processing steps: MRI sequences are at isotropic 1 mm^3 voxel resolution, with T1, T1ce, T2 and FLAIR MRI modes aligned to the same anatomical reference frame. Voxel-to-voxel matching between different sequences was provided. To facilitate tumor analysis, areas such as the skull and eyes have been removed, leaving only brain tissue. By applying histogram-based normalization, brightness and contrast changes in MRI images were reduced.

The images in the BraTS 2019 dataset contain pixels from four different classes: necrotic and non-enhancing tumor, contrast-enhancing tumor, peritumoral edema, and background. The tumor core consists of the growing tumor, necrotic and non-enhancing tumor, and the entire tumor consists of the growing tumor, peritumoral edema, and non-enhancing necrotic tumor. A growing tumor is tumor tissue containing viable and malignant cells and continues to spread to other parts of the brain. Peritumoral edema, on the other hand, is the excessive fluid accumulation that usually occurs in response to a malignant tumor, creating a rapidly growing area around the tumor. As this area of edema expands, it can compress other parts of the brain, causing damage. A necrotic and non-enhancing tumor area is a significant indicator that the lesion is malignant and corresponds to necrotic tissue.

In this study, operations were performed on the entire tumor and the growing tumor during the training and evaluation phases. The necrotic-non-enhancing tumor and enhancing tumor were considered positive; the peritumoral edema area and the background were considered negative. While the images were considered as training and testing datasets, 5-fold cross-validation was applied. For this purpose, the data was divided into parts, 80% of which was training and 20% was test data. Then, the results obtained after the run for each fold (part) was averaged. The proposed models were trained with labelled pixels of the training data. The classification results of each pixel in the test dataset were transferred to a confusion matrix, and four different statistical evaluation metrics were calculated based on this matrix. Thus, detailed analysis of model performance was possible.

To evaluate the performance of the models trained using the BraTS 2019 dataset, additional tests were performed with an independent dataset, Figshare [14]. The Figshare dataset consists of 1,372 MRI images. The images are 256x256 pixels in size, and only size standardization and contrast adjustments were made. Here, the Figshare dataset differs from BraTS 2019 because it is preprocessed. The performance metrics Dice Coefficient, Sensitivity, Specifity, and Jaccard Index used in BraTS 2019 were also used to evaluate the Figshare dataset. The results are presented in Table 2 for comparison with the BraTS 2019 results.

2.1. VGG-16

Earlier VGG neural network models, such as AlexNet, focus on smaller window sizes and strides in the first convolutional layer [27]. VGG-16, on the other hand, is a CNN model and addresses another crucial aspect of CNNs, depth. An example VGG model is shown in Figure 1. This VGG treats each image as 224x224 pixels RGB, thus keeping the input image size consistent. The convolution layers in VGG use the smallest possible space size. Furthermore, existing 1x1 convolution filters perform a linear transformation on the input and are then activated with a ReLU unit. The convolution operation preserves spatial resolution by applying 1-pixel steps. The VGG architecture consists of three fully connected layers. The first two layers have 4096 channels, and the third layer has 1000 channels. Since all hidden layers of VGG use ReLU, training time is reduced [8].

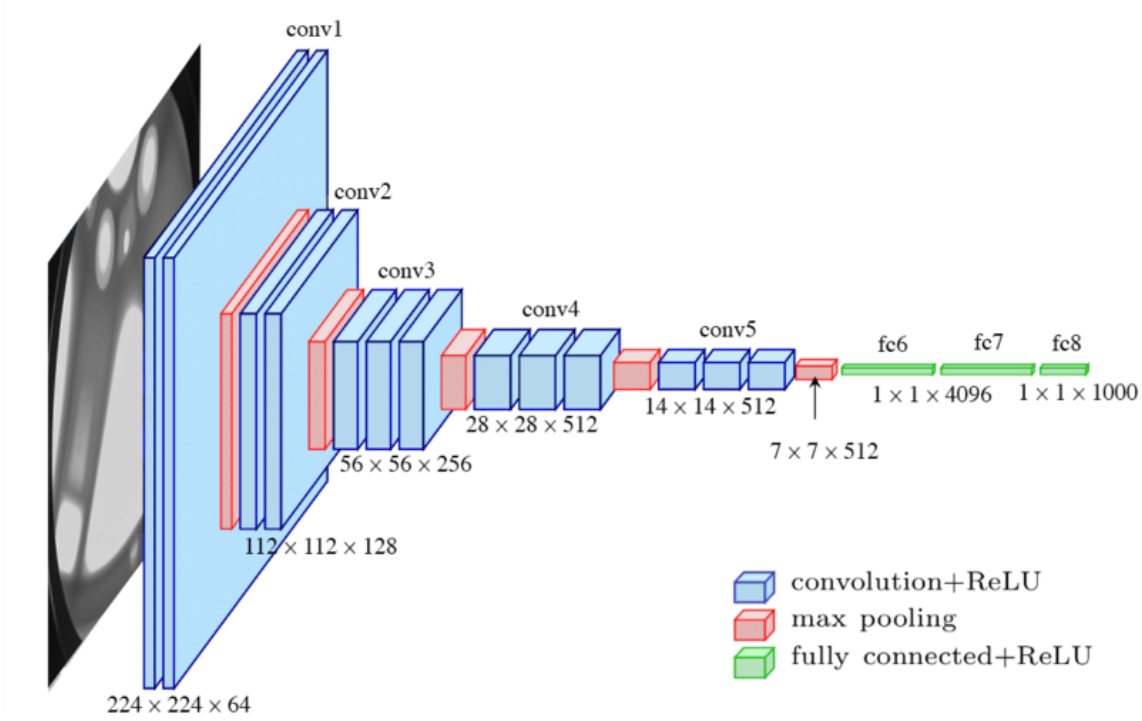

Figure 1. VGG-16 architecture [28]

2.2. ResNet50

By adding additional layers to Deep Neural Networks, better accuracy and performance are achieved in complex problems. By adding more layers, the network is intended to learn increasingly complex layers. Different layers can be considered to recognize edges, textures, and objects in an image. However, in CNNs there is a maximum threshold for depth [9]. ResNet has made a positive contribution to the problem of training Deep Networks. An example of a residual block is shown in Fig. 2. There is a

connection that directly skips layers, called a Jumping Connection. This connection forms the core of the blocks and provides a different output [15].

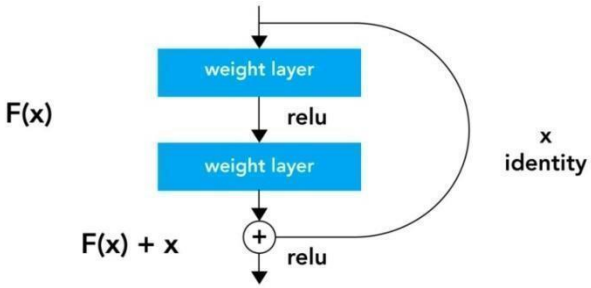

Figure 2. Residual block diagram [15]

As shown in Figure 3, each 3x3 convolution layer in the residual block is followed by a batch normalization layer and then a ReLU layer. After this, the convolution operations are skipped, and the input is directly added to the output before the final ReLU function. At this point, the output of the two convolution layers is expected to have the same structure as the input, allowing for summation [15]. In this study, the first four layers are used for the U-Net + ResNet model.

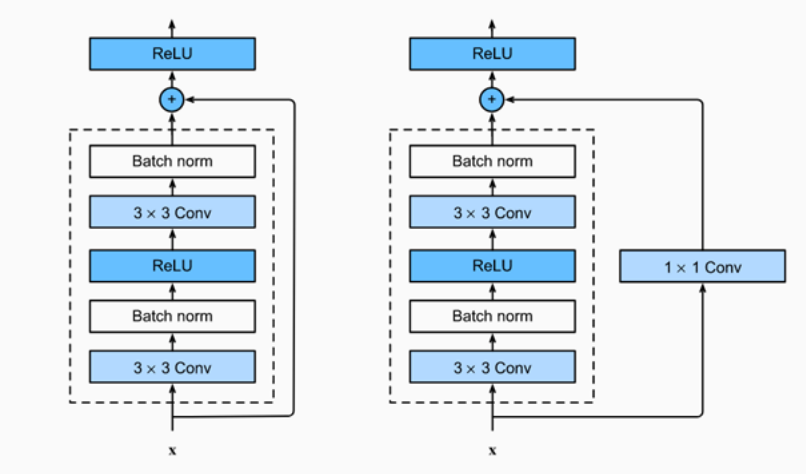

Figure 3. ResNet block and ResNet block with 1×1 convolution [16]

The input size may differ from the size of the output, which can occur with pooling layers and convolutional layers in between. In such a problem, it is possible to fill the jump link with extra zero entries without using an additional parameter to increase its size. Another way is to add an additional 1x1 layer when the number of channels is desired to be changed while adjusting the input to the desired shape during the summation process [15]. For this purpose, additional parameters are utilized. Initially inspired by VGG-19, the ResNet architecture has a flat 34-layer network structure. Later, skip connections were added to this structure. Thus, the architecture became known as "ResNet" [9].

2.3. U-net

U-Net is a CNN approach developed by Ronneberger et al. [6] to provide more accurate segmentation of biomedical images. While traditional CNNs typically treat the output of an image as a single class label, in fields like defense and biomedical applications where very precise measurements are required, information is required for each pixel. In other words, each pixel requires a separate class

label [15]. In addition, large data sets are required in CNNs. Accessing and storing large data sets is not always sufficient, and the time and human energy needed to process this data is not sufficient. In their study, Ronneberger et al. [6] introduced the U-Net model, which produces more accurate results, based on a fully convolutional network structure [18]. The U-Net architecture provides clearer segmentations while also allowing for effective results to be obtained with fewer training images.

In a study by Shelhamer et al. [18], a joint contraction network was supplemented with successive layers and upsampling operators were used instead of pooling operations. These layers allow for increased resolution of the output. In this way, a successful convolution layer can learn to produce precise outputs. The U-Net structure also allows the network to transfer context information to higher-resolution layers by incorporating multiple feature channels during the sampling phase. The widening path is more or less symmetrical to the narrowing part. As seen in Fig. 4, it gives a U-shaped appearance and takes its name from this appearance. Here, the network uses only the valid portion of each convolution and performs this without fully connected layers [18]. The missing context is estimated by projecting the input image. This also makes it possible to estimate pixels in the image's border region. Applying the network to large images with this strategy ensures that the resolution is not limited by GPU memory.

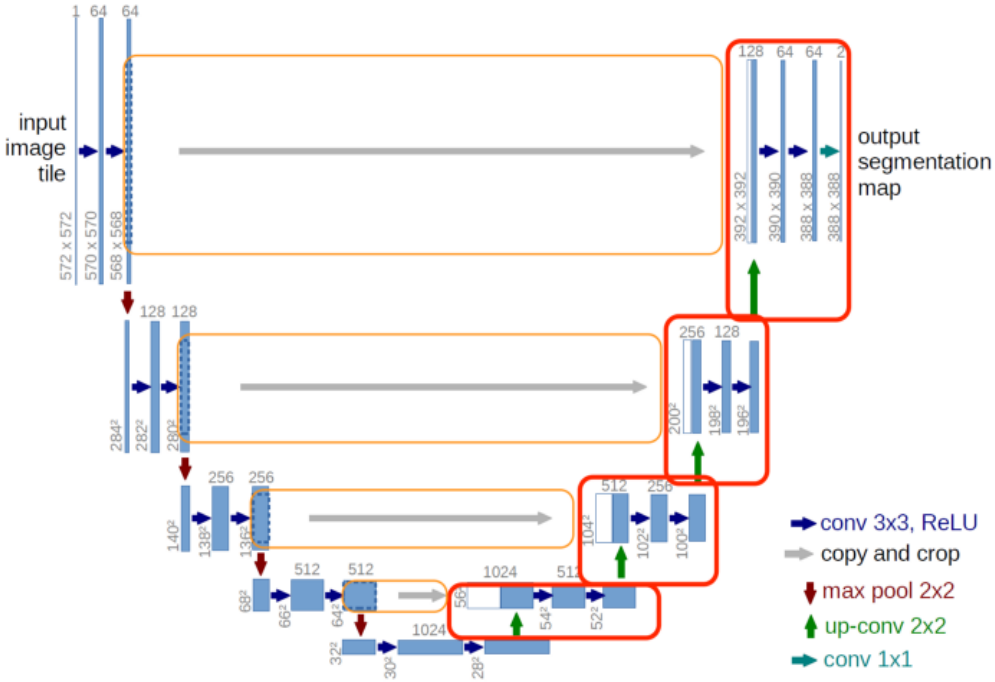

Figure 4. Classic U-Net model [6]

The U-Net architecture consists of two basic parts: a narrowing path on the left and an expanding path on the right (Fig. 4). Contracting path: This is a typical convolutional network structure, with each stage consisting of a 3x3 convolution. This is followed by downsampling with the ReLU activation function and 2x2 max pooling. The number of feature channels is doubled at each stage with the downsampling process. Expanding path: This is an upsampling structure implemented with a 2x2 "upconvolution." It consists of two 3x3 convolutions, each with a ReLU activation. These layers are concatenated with feature maps appropriately cropped from the contracting path. During upsampling, the number of "feature channels" is halved. Because edge pixels are lost during each convolution, cropping is applied to prevent this loss. A 1x1 convolution is used in the final layer to transform each 64-dimensional feature vector into the desired number of classes. This network consists of 23 layers [6]. Figures 5 and 6 illustrate the two main parts of the U-Net architecture: the contraction and expansion paths. The distinguishing feature of this architecture is the second half, which specifically includes the

expansion path. The U-Net architecture lacks fully connected layers; only convolution layers are used, and these layers operate with the ReLU activation function [6].

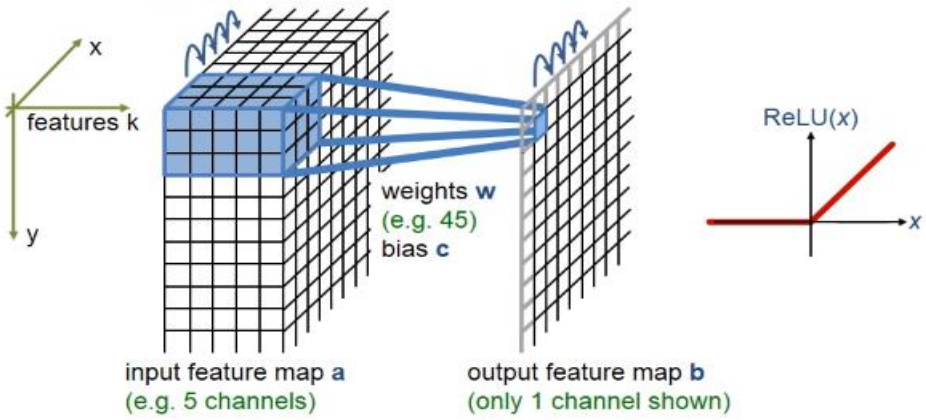

Figure 5. 3x3 convolution and ReLU representation [6]

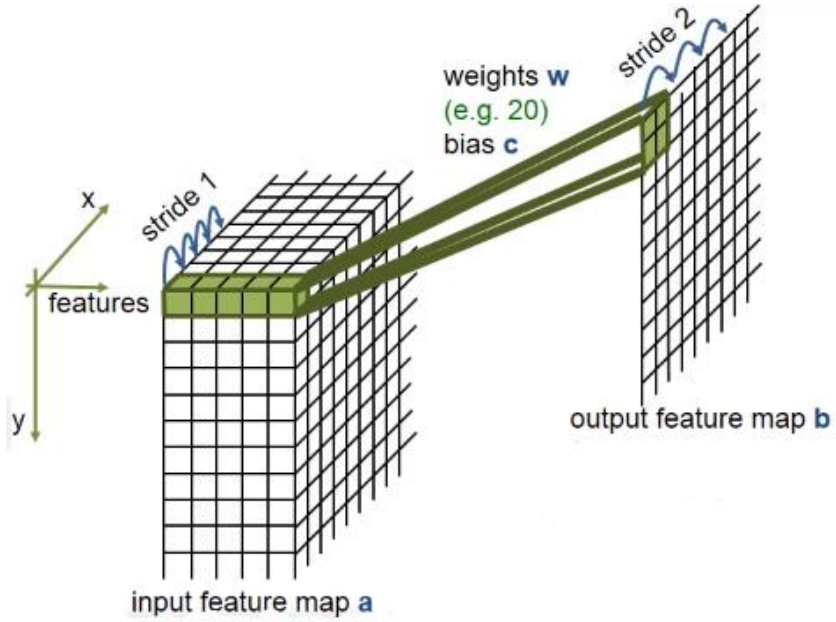

Figure 6. 2x2 Up-convolution representation [6]

2.4. U-net ++

Zhou et al. [12] introduced a new segmentation architecture, U-Net++, based on nested and dense skip connections to achieve more precise segmentation in medical images. In U-Net++, feature maps from the encoder network are progressively enriched before being combined with the decoder network. This allows the network to more effectively capture the details of foreground objects. The learning performance of the network is also improved when the feature maps from the encoder and decoder networks are semantically similar [12].

U-Net++ differs from the classical U-Net by reducing the semantic gap between the encoder and decoder feature maps, thanks to the convolution layers included in the skip paths. Furthermore, dense

skip connections improve gradient flow over time and enable model pruning with a deep control mechanism, improving overall model performance.

UNet++ consists of an encoder and decoder structure connected via nested dense convolutional blocks (Figure 7). This architecture utilizes the deep control method introduced by Lee et al. [20]. The model can operate in two different modes: accuracy mode and fast mode. In accuracy mode, the outputs from all segmentation branches are averaged. In fast mode, the output from only one segmentation branch is used. When UNet++ is trained with deep control, it can be made more efficient by pruning during inference.

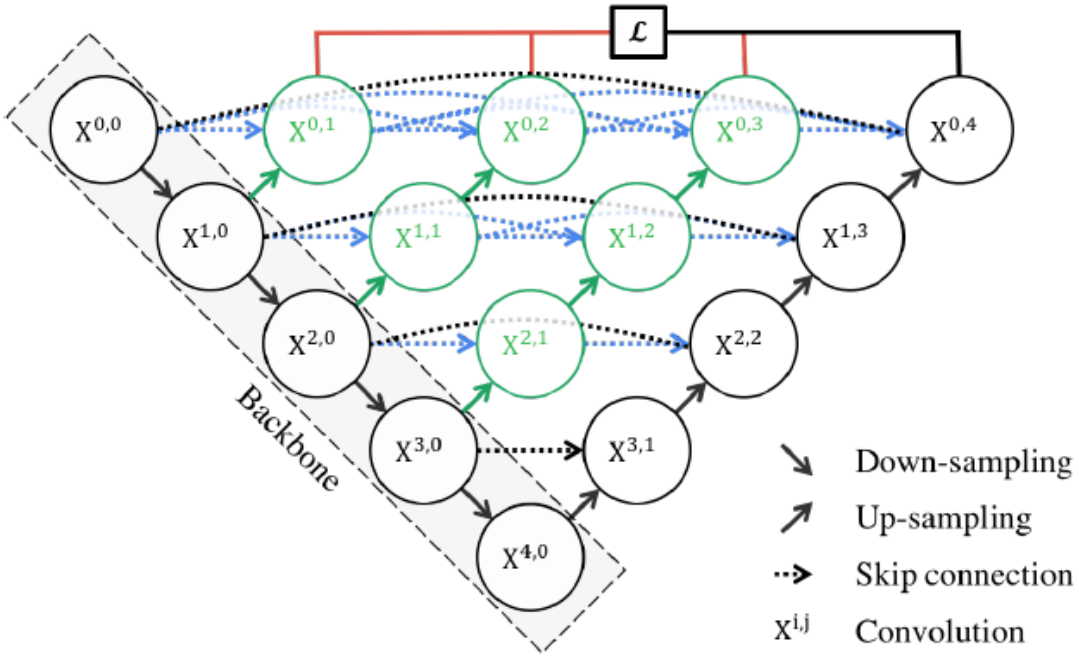

Figure 7. U-Net++ model [12]

2.5. Training Protocol and Hyperparameters

All proposed models—Classic U-Net, U-Net+VGG-16, U-Net+ResNet50, and U-Net++—have been standardized as much as possible to allow for comparable evaluation. However, due to limitations arising from architectural differences, some parameters have been adjusted.

The "Adaptive Moment Estimation" optimization algorithm was used to train all models: This algorithm, with its adaptive learning rate, provided stable convergence across different architectures. A constant initial learning rate of 1e-4 (0.0001) was set for all experiments. This value is within a range commonly used for stable training to prevent "overfitting" in sensitive operations such as medical image segmentation.

The batch size was determined considering GPU memory constraints and model stability. For Classic U-Net & U-Net++ (relatively lighter models), a batch size of 16 was used; The batch size for U-Net+ResNet50 & U-Net+VGG-16 (deep and heavy hybrid models) was reduced to 8. This was done to minimize memory consumption and allow for training. All models were trained for 400 epochs for more consistent processing to observe and compare convergence behaviour. Dice loss was used to optimize segmentation performance and effectively address tumor-background class imbalance.

To improve the model's generalization ability and reduce overtraining, a basic "data augmentation" strategy involving random horizontal/vertical flips, scaling, and angular rotations was applied during the training phase. 50% of the 284 brain images in BraTS2019 and 1,372 brain images in Figshare were horizontally/vertically flipped; each image was rotated by an angle of ±10°, and each image was scaled

between 90% and 110%. These operations increased the diversity of the dataset while preserving the basic geometric structure of the images.

2.6. Performance Evaluation

To evaluate the performance of the classical U-Net, U-Net + VGG-16, U-Net + ResNet50, and U-Net++ networks, parameters such as dice coefficient, sensitivity, specificity, and Jaccard Index were calculated based on the confusion matrix data. The confusion matrix [21] is a table used to evaluate the performance of a classification model using the actual values in a test dataset. This concept is illustrated in Table 1.

Table 1. Confusion matrix

		REAL VALUE	
		POSITIVE	NEGATIVE
ESTIMATED VALUE	*POSITIVE*	TP True-Positive	FP False-Positive
	NEGATIVE	FN False-Negative	TN True-Negative

2.7. Accuracy Rate

Accuracy rate is a measure of the accuracy of the prediction made by the classifier. It is formulated as in Eq. 1.

$$Accuracy\ rate = \frac{(TP + TN)}{Total} \tag{1}$$

2.8. Dice Coefficient

It is a statistical method used to calculate the similarity of two samples. When the positive elements in the result match the ground truth, this result is divided by the total number of positive elements in both samples. This process is expressed by the formula in Equation 2.

$$Dice\ coefficient = \frac{2|A \cap B|}{|A|+|B|} = \frac{2(TP)}{(FP)+(FN)+2(TP)} \tag{2}$$

2.9 Sensitivity

Sensitivity indicates the proportion of positive data points that are considered positive out of all positive data. It is expressed as in Eq. 3.

$$Sensitivity = \frac{TP}{TP+FN} \tag{3}$$

2.10. Specificity

Specificity measures the ability of the model to correctly identify negative samples. It is expressed as in Eq. 4.

$$Specificity = \frac{TN}{TN+FP} \tag{4}$$

2.11. Jaccard Index

It is one of the pixel-based criteria used when evaluating segmentation performance. The overlapping ratio between the target matrix and the predicted matrix is taken into account. It is formulated as in Eq. 5.

$$Jaccard\ index = \frac{|A \cap B|}{|A \cup B|} = \frac{TP}{(TP+FP+FN)} \tag{5}$$

3. RESULTS AND DISCUSSION

The Classic U-Net model was first applied to the dataset. In the study, 80% of the original tumor MRI images in the BraTS dataset were used as training data, and the remaining portion was used as testing data (5-fold cross-validation). The model was operated for 400 epochs. In general, it was observed that the results converged after 150 epochs. In particular, no significant increase was observed after this point in the accuracy graph. Accuracy, Loss, Dice coefficient and Jaccard index values are shown in Fig. 8. Additionally, a sample HGG MR image in the dataset, the original mask, and the values predicted with the Classic U-Net are shown in Fig. 9.

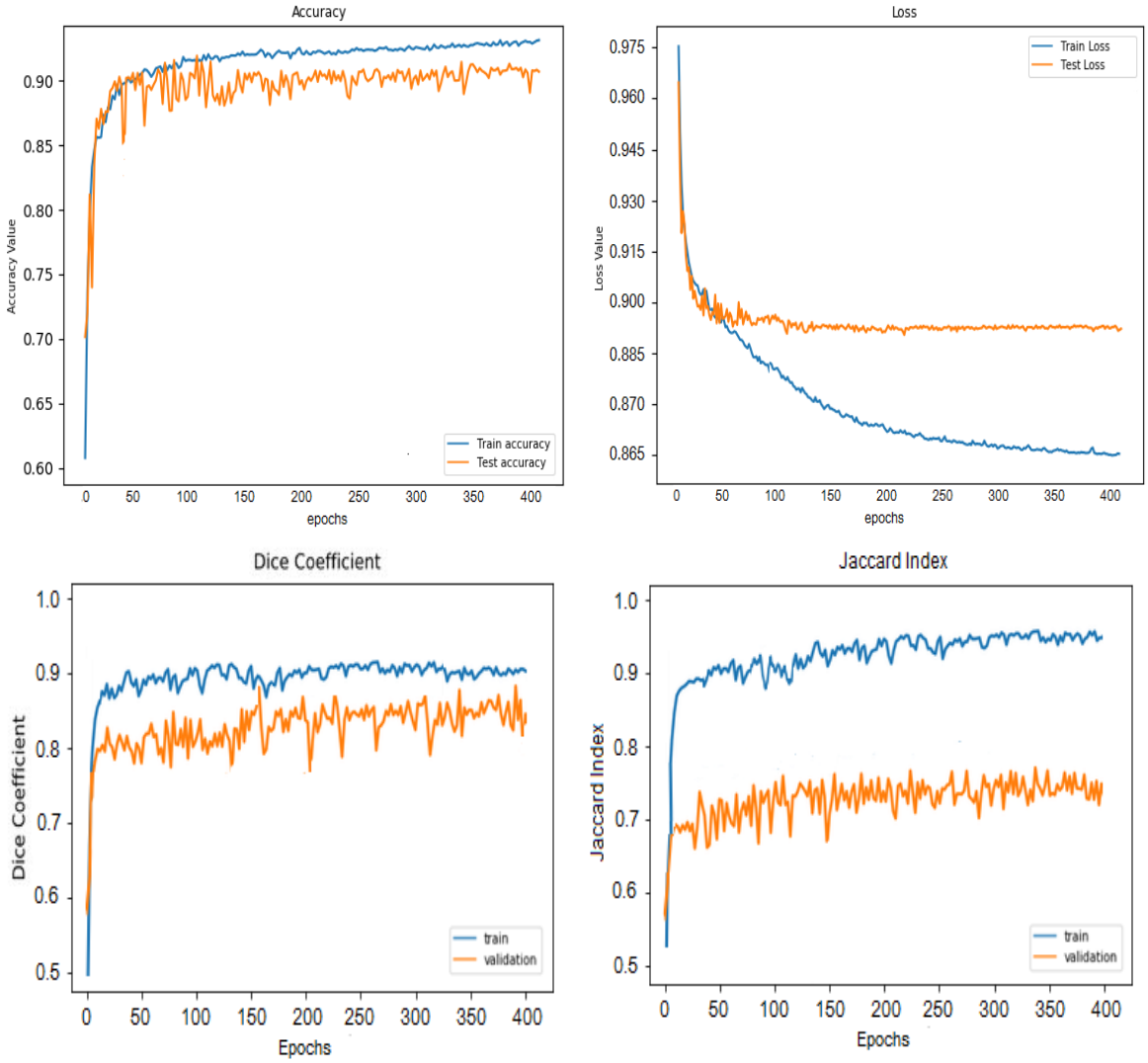

Figure 8. U-Net performance graphs of segmented tumors

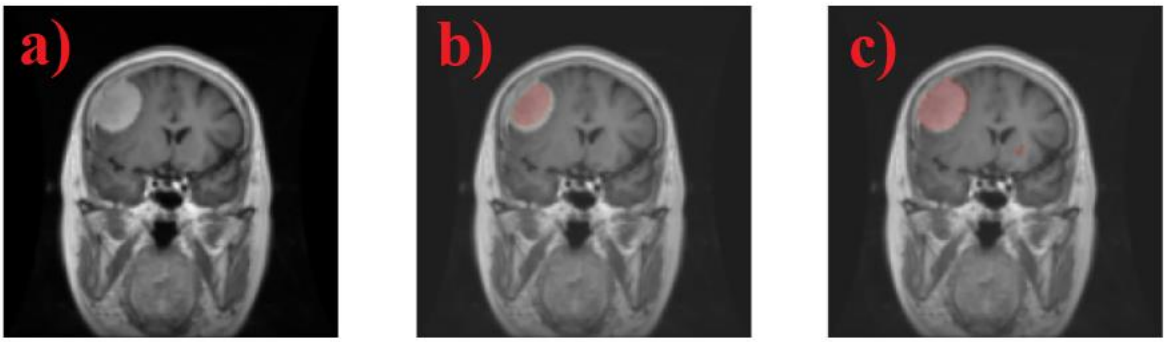

Figure 9. HGG MR image for U-Net, a): MRI image, b): original mask, c): predicted mask

The second model considered is the U-Net + VGG-16 architecture, Partial VGG-16 integration. Here, the Block3, Block4 and Block5 outputs of the VGG-16 model are used as feature extractors for segmentation. By removing the first two blocks (Block1 and Block2) (not used as skip connections), computational complexity is reduced and low-level (edge, texture) information is prevented from causing over-learning [22], [23]. The bridge layer (Bottleneck) is converted into a feature vector by processing the last convolutional block output of VGG-16 (Block5) with 1×1 convolution and batch normalization. Feature maps are brought to the original resolution by transposed convolution (up-sampling). The Block3 and Block4 outputs from the middle and deep layers of VGG-16 are combined with the decoder layers of U-Net via skip connections. 3×3 convolution + ReLU activation + batch norm is applied in each layer [11], [12]. The performance results obtained with this model are given in Fig. 10. An example HGG MR image in the dataset as the output of the system created with this hybrid approach (original mask and values estimated with U-Net + VGG-16) is presented in Fig. 11. As another hybrid approach, U-Net + ResNet50 architecture is proposed. ResNet50 was used in the encoder (feature extractor) part of U-Net. In this integration, multi-scale feature maps taken from the middle layers (3rd block, 512 feature channels) and deep layers (4th block, 1024 feature channels) of ResNet50 were transferred to the decoder part via 'skip connections' [9]. In this way, segmentation accuracy is optimized by using low-level edge and texture information together with high-level semantic features. The deepest layer output of ResNet50 (5th block, 2048 feature channels) constitutes the bottleneck layer of the model. In the decoder section, the transposed convolution layers (Conv2DTranspose) and spatial resolution are gradually increased, and the skip connections from the encoder are combined with the "concatenate" operation to preserve local details. This hybrid approach provides an effective transfer learning strategy, especially for limited medical image data, and aims to minimize the Dice loss through multi-scale feature fusion [9], [12], [25], [26].

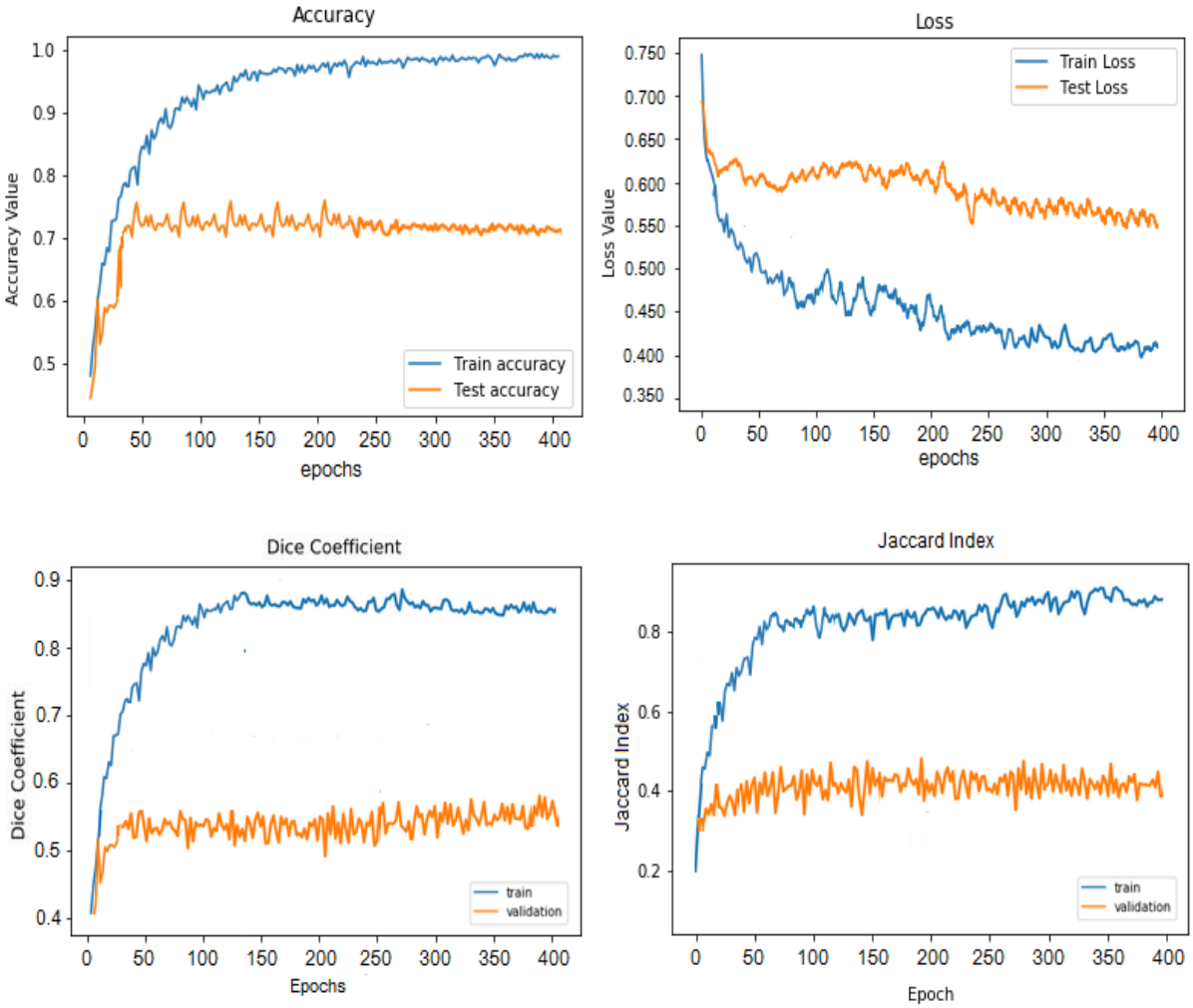

Figure 10. U-Net + VGG-16 performance graphs of segmented tumors

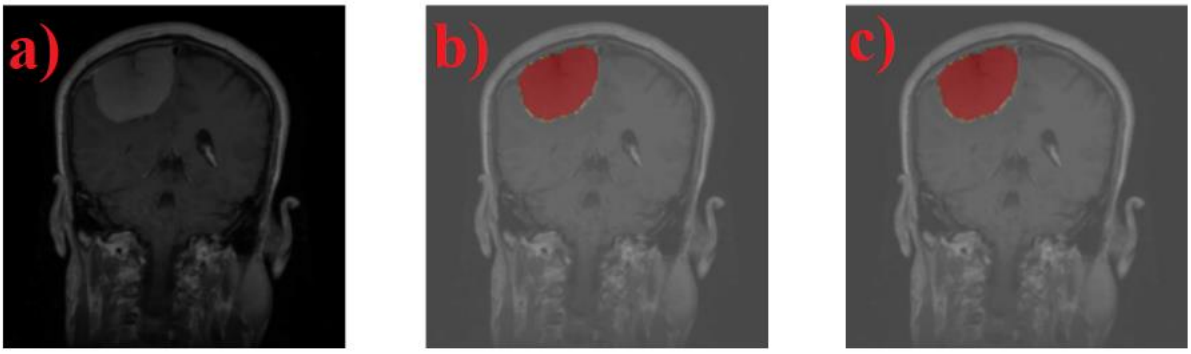

Figure 11. HGG MR image for U-Net + VGG-16, a): MRI image, b): original mask, c): predicted mask

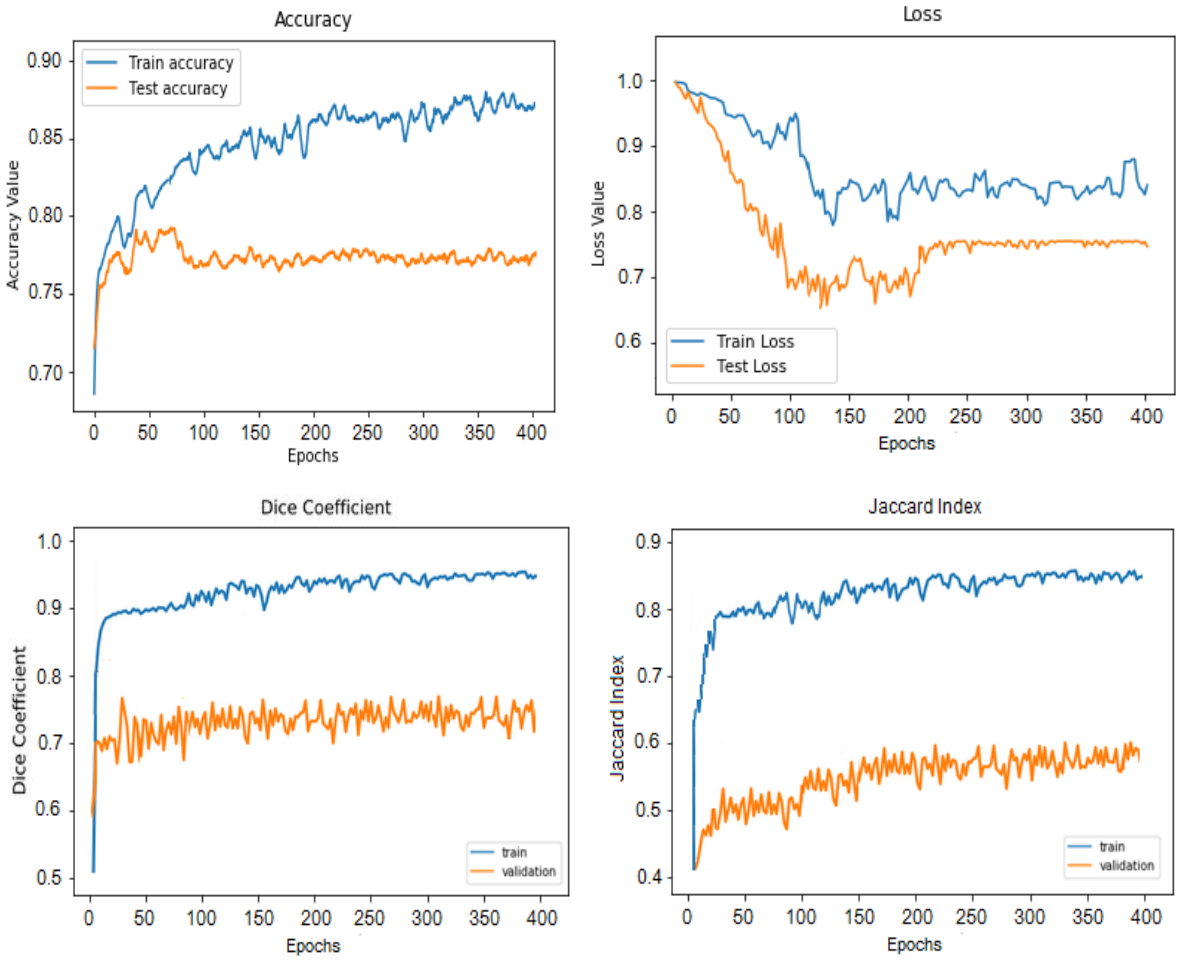

Figure 12. U-Net + ResNet50 performance graphs of segmented tumors

The performance results obtained with this model are given in Fig. 12. Fig. 13 shows the resulting images (original mask and prediction made with U-Net + ResNet50) obtained with this hybrid approach for a sample HGG MR image in the dataset.

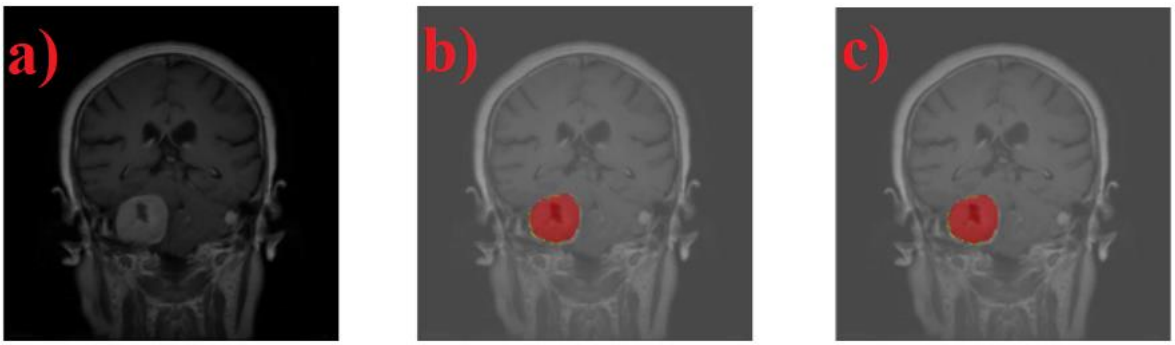

Figure 13. HGG MR image for U-Net + ResNet50, a): MRI image, b): original mask, c): predicted mask

Finally, the U-Net++ architecture is discussed. The performance results obtained are given in Fig. 14. An example HGG MR image in the dataset, the original mask, and the values predicted by U-Net++ are shown in Fig.15. For some images in the Figshare dataset [14], the original MRI image, the original tumor labels in these images, and the prediction results obtained with the proposed models (Classic U-

Net, U-Net+VGG-16, U-Net+ResNet-50, and U-Net++ architectures) are presented in Fig. 16, 17, 18, and 19.

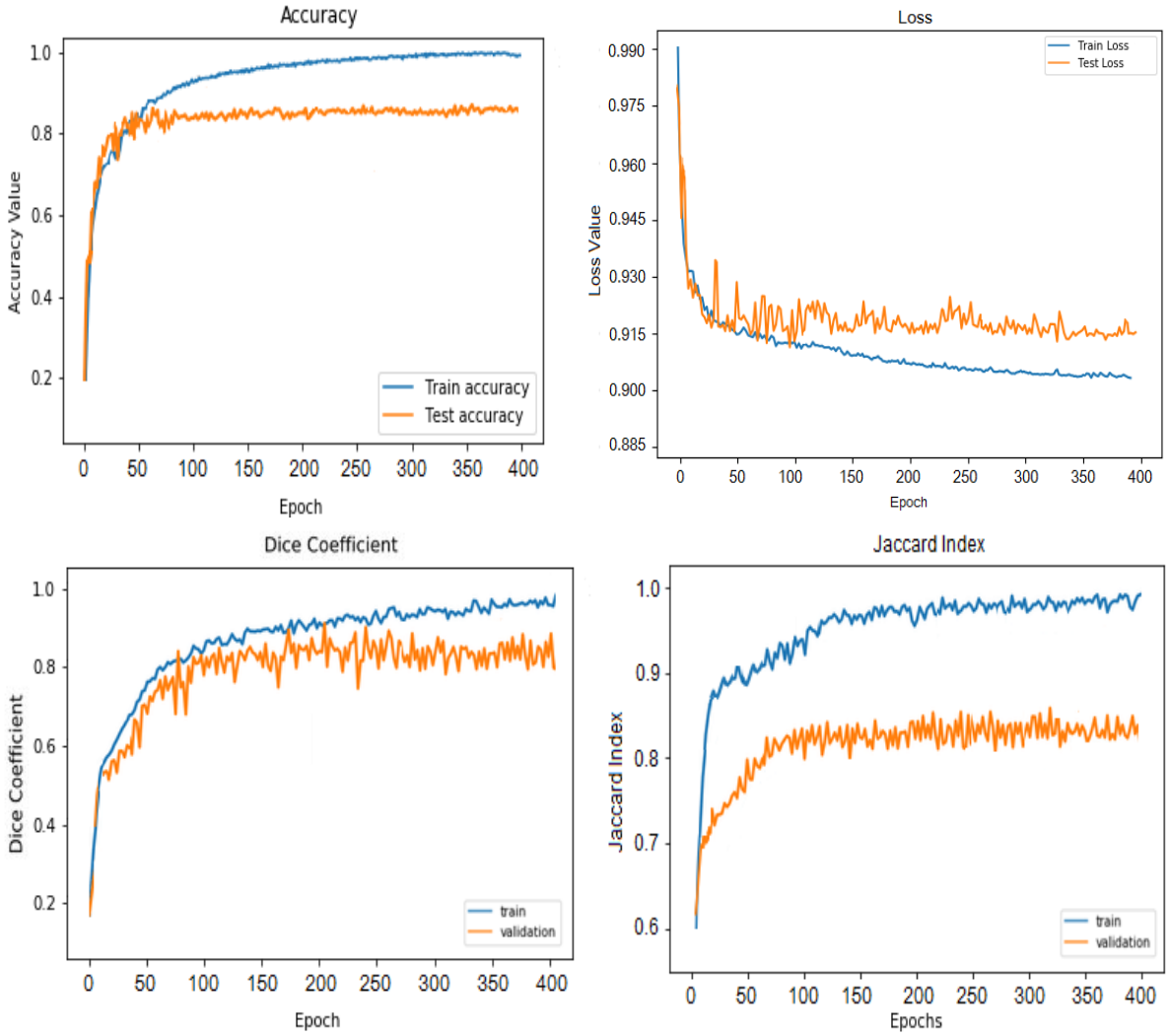

Figure 14. U-Net++ performance graphs of segmented tumors

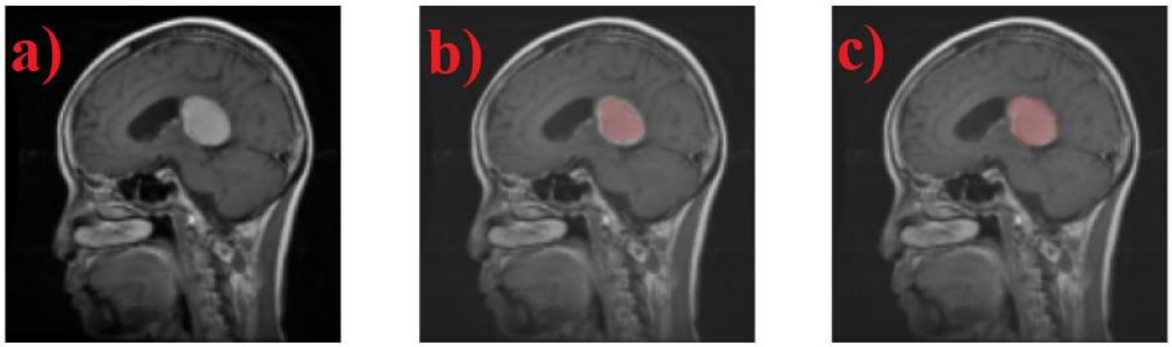

Figure 15. HGG MR image for U-Net++, a): MRI image, b): original mask, c): predicted mask

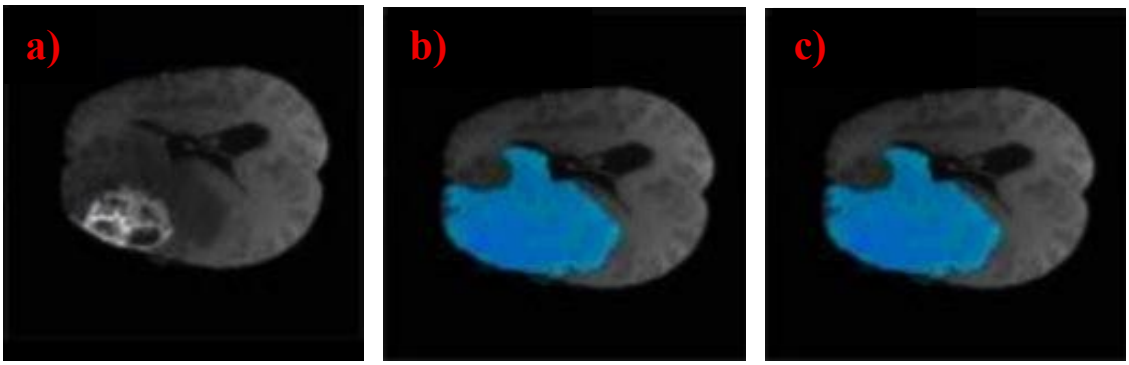

Figure 16. MR image for U-Net, a): MRI image, b): original mask, c): predicted mask

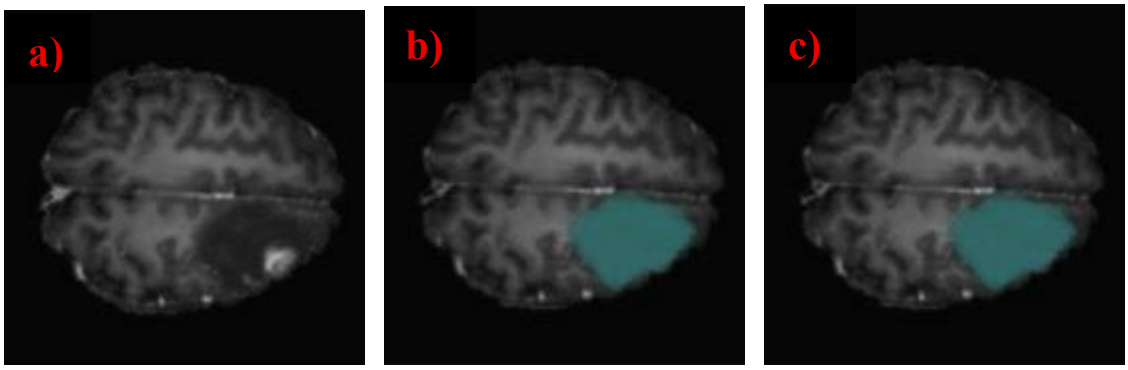

Figure 17. MR image for U-Net+VGG-16, a): MRI image, b): original mask, c): predicted mask

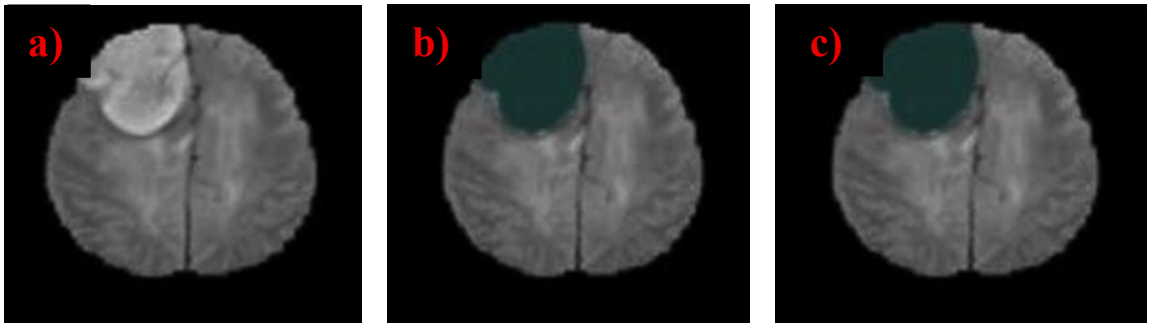

Figure 18. MR image for U-Net+ResNet50, a): MRI image, b): original mask, c): predicted mask

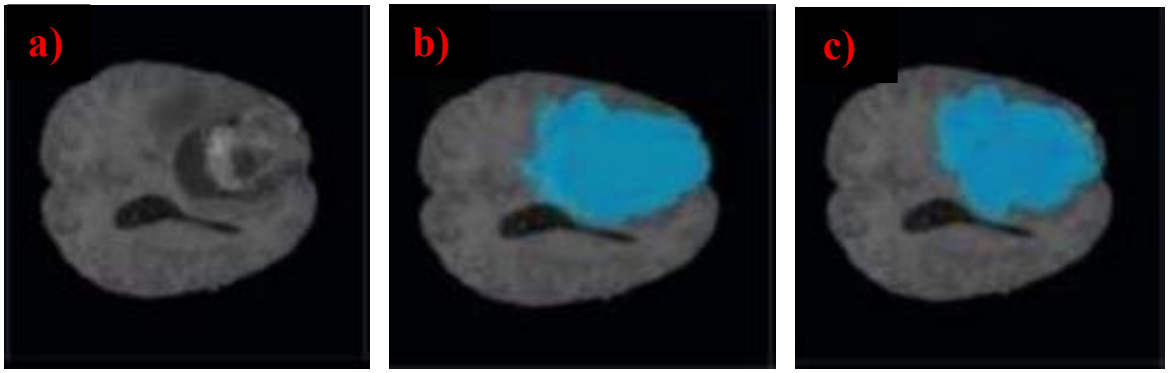

Figure 19. MR image for U-Net++, a): MRI image, b): original mask, c): predicted mask

Table 2 shows the Dice Coefficient, Sensitivity, Specificity and Jaccard Index results of various architectures applied to the BraTS 2019 dataset. When the studies in the literature are examined, it is seen that the best success is achieved with the U-Net++ architecture.

Table 2. Performance metrics obtained from different architectures of brain tumor

Applications	Dataset	Dice Coefficient	Sensitivity	Specificity	Jaccard Index
Classic U-Net	BraTS 2019	0.879	0.844	0.957	0.784
	Figshare	0.910	0.895	0.968	0.835
U-Net+ResNet	BraTS 2019	0.697	0.659	0.862	0.535
	Figshare	0.750	0.720	0.890	0.600
U-Net+VGG-16	BraTS 2019	0.585	0.550	0.800	0.414
	Figshare	0.650	0.630	0.850	0.480
U-Net++	BraTS 2019	**0.920**	**0.890**	**0.973**	**0.850**
	Figshare	0.905	0.880	0.960	0.827

In this study, the performance of the hybrid models U-Net+ResNet50 (Dice: 0.697) and U-Net+VGG-16 (Dice: 0.585) was significantly below that of the other models. The reasons for this low performance are: 1. 2D Encoder – 3D data mismatch. The BraTS 2019 dataset consists of 3D MRI volumes with dimensions of 240x240x155, while encoders such as VGG-16 and ResNet50 are 2D convolutional neural networks trained on natural images (ImageNet). In these hybrid models, although each slice of 3D MRI data is processed as a 2D image, the encoder layers cannot capture the volumetric context and three-dimensional structural relationships between these 2D slices. This results in the loss of 3D contextual information, which is crucial for understanding the three-dimensional morphology and neighborhood relationships of tumors, and this significantly reduces segmentation accuracy. 2. Unoptimized transfer learning: The theoretical power of hybrid models emerges from the rich feature maps provided by pre-trained encoders. However, in this study, the fine-tuning required for these encoders (VGG-16 and ResNet50) to adapt to the specific structure of the medical images in the BraTS and Figshare datasets could not be implemented systematically and adequately due to computational resources and time constraints. Because the learning rates and freezing strategies of the encoder layers were not optimized, the general features learned from ImageNet could not be effectively combined with the texture and density distribution of the medical images. This results in inefficiency in feature extraction and low segmentation quality. 3. Lack of hyperparameter optimization: Comparing the hybrid architectures to Classic U-Net or U-Net++, we observe that they have more complex structures. Therefore, more precise hyperparameter optimization is required in hybrid architectures. Critical hyperparameters were fixed without an exhaustive search for these models. In particular, the failure to select loss functions to address the class imbalance (tumor-body) problem frequently encountered in medical images may have limited our model's ability to detect small tumor regions. 4. Architecture and skip connection mismatch: While U-Net++ provides successful results with its dense and nested skip connection structure, which reduces the semantic gap between the encoder and decoder, in the U-Net+VGG-16 and U-Net+ResNet50 models, we observe that the skip connections between the feature maps obtained from the pre-trained encoders and the U-Net decoder are direct and unoptimized. We observe a mismatch between the feature size and semantic level in the encoder output and the structure expected by the decoder, rendering information flow inefficient.

In addition to the poor performance of hybrid models, Table 2 reveals dataset-dependent behavior among the remaining architectures. In the Figshare dataset, the classic U-Net outperforms more complex variants, achieving the highest Dice score (0.910). In the BraTS 2019 dataset, U-Net++ demonstrates superior performance (Dice: 0.920). The Figshare brain tumor dataset includes MR images of three different tumor types (glioma, meningioma, and pituitary tumors) and, unlike BraTS, does not include multi-parameter modalities. Due to this limited modality and tumor diversity, the dataset may exhibit relatively less complex imaging variability compared to the multi-modal BraTS dataset. On the other hand, the BraTS dataset includes multi-parameter MR scans from numerous centers with high within-patient variability and complex glioma morphology, which can increase the difficulty of segmentation tasks. U-Net++'s dense and nested link design effectively reduces the semantic gap between encoder and

decoder features, enabling improved multi-scale feature fusion. This architectural advantage allows U-Net++ to better model the complex spatial and contextual relationships present in BraTS 2019. It also improves segmentation accuracy compared to simpler architectures.

Table 3. Comparison of the obtained results with some studies in the literature

Study	Architecture	Dataset	Dice coefficient
Jiang et al. [17]	Two-stage U-Net	BraTS 2019	0.88
Liu et al. [19]	LR-Net	BraTS 2019	0.80
Jena et al. [24]	Depth-reduced U-Net	BraTS 2019	0.88
Jiang et al. [29]	SwinBTS	BraTS 2019	0.89
Liu et al. [30]	CSWin-Unet	BraTS 2019	0.90
Rasool et al. [31]	TransRes-Unet	BraTS 2019	0.91
Current study	**U-Net++**	**BraTS 2019**	**0.92**

Table 3 presents a comparison between recent studies conducted on the BraTS 2019 dataset and the performance of the proposed U-Net++ model. The table compares the WT (whole tumor) values reported in the literature with those obtained by the U-Net++ model. When the table is examined, it is seen that the U-Net++ architecture proposed in the current study outperforms all other architectures in the list with its high segmentation accuracy (Dice coefficient of 0.92). While the results (0.88 and 0.80) obtained by Jiang et al. [17] and Liu et al. [19] with two-stage U-Net and LR-Net architectures, respectively, are consistent with this study. The U-Net model proposed by Jena et al. [24] (Dice score of 0.89) confirms the effectiveness of U-Net-based architectures. Our study shows that the dense skip connection structure of U-Net++ is more successful in reducing the semantic gap compared to such variations.

In the study, Jiang et al. [29] achieved a score of 0.89 with the SwinBTS model; Liu et al. [30] achieved a score of 0.90 with the CSWin-Unet model; Rasool et al. [31] achieved a score of 0.91 with the TransRes-Unet model. These results summarize the tumor segmentation performance of the mentioned methods on the BraTS 2019 dataset and demonstrate that UNet++, the subject of this study, performs better than the other methods.

4. CONCLUSIONS

In this study, four different deep learning architectures were examined for brain tumor segmentation and it was shown that U-Net++, with its dense skip connection structure, exhibited the highest performance, while Classic U-Net also performed well. Hybrid models yielded lower performance compared to other architectures due to 2D-3D mismatch and unoptimized transfer learning. For future studies, validation of these evaluated models with different, multi-center data is recommended to enhance their functionality for use in hospitals or clinics. Furthermore, integrating them with different imaging modalities such as PET and CT, beyond the current MRI-based approach, could improve tumor characterization. Adding spatial and channel attention structures to the U-Net++ framework can strengthen the model's ability to focus on tumor regions and provide architectural improvements. For performance optimization, it is recommended to use combined loss functions such as Dice + Focal Loss or Tversky Loss, which handle class inconsistency more effectively, and ultimately, to transition to volumetric architectures such as 3D U-Net++ to capture the three-dimensional context information lost in 2D slices. These trends will enable future models to perform good segmentation, thus effectively increasing their clinical applicability.

Declaration of Ethical Standards

The authors declare to comply with all ethical guidelines, including authorship, citation, data reporting, and original research publication.

Credit Authorship Contribution Statement

The authors declare that they have made equal contributions.

Declaration of Competing Interest

The authors declare that they have no conflict of interest.

Funding / Acknowledgements

No funding was received for conducting this study.

Data Availability

Data will be made available on request.

Use of Artificial Intelligence

Generative artificial intelligence (AI) tools were not used to generate scientific content, research data, interpretations, or analyses. The authors acknowledge that the use of AI tools is limited to language editing and readability improvements.

REFERENCES

[1] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts and P. Walter, "Molecular biology of the cell (6th ed.)", *Garland Science*, 2014.
[2] R. Paudyal, A. D. Shah, O. Akin, R. K. G. Do, A. S. Konar, V. Hatzoglou, U. Mahmood, N. Lee, R. J. Wong, S. Banerjee, J. Shin, H. Veeraraghavan and A. Shukla-Dave, "Artificial intelligence in CT and MR imaging for oncological applications", *Cancers*, vol. 15, no. 9, pp. 2573, 2023, doi: https://doi.org/10.3390/cancers15092573.
[3] C. Liao and L. Kong "A comprehensive survey of semantic segmentation based on computer vision", *Proc. SPIE 13486, Fourth International Conference on Computer Vision, Application, and Algorithm (CVAA 2024)*, 1348612, 2025, doi: https://doi.org/10.1117/12.3055943.
[4] A. I. Khan, J. L. Shah and M. M. Bhat, "Coronet: A deep neural network for detection and diagnosis of COVID-19 from chest X-ray images", *Computer Methods and Programs in Biomedicine*, vol. 196, pp. 105581, 2020, doi: https://doi.org/10.1016/j.cmpb.2020.105581.
[5] A. Esteva, K. Chou, S. Yeung, N. Naik, A. Madani, A. Mottaghi, Y. Liu, E. Topol, J. Dean and R. Socher, "Deep learning-enabled medical computer vision", *NPJ Digital Medicine*, vol. 4, no. 1, 1-9, 2021, doi: https://doi.org/10.1038/s41746-020-00376-2.
[6] O. Ronneberger, P. Fischer and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation", *In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015, Lecture Notes in Computer Science, 9351. Springer, Cham*., 2015, doi: https://doi.org/10.1007/978-3-319-24574-4_28.
[7] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation", *Lecture Notes in Computer Science*, vol. 9901, 424-432, 2016, doi: https://doi.org/10.1007/978-3-319-46723-8_49.

[8] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", *Computer Vision and Pattern Recognition*, pp. 1-14, 2015, doi: https://doi.org/10.48550/arXiv.1409.1556.
[9] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition", *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA*, pp. 770-778, 2016, doi: https://doi.org/10.1109/CVPR.2016.90.
[10] H. Dong, G. Yang, F. Liu, Y. Mo and Y. Guo, "Automatic brain tumor segmentation using U-Net based fully convolutional networks", *Computer Vision and Pattern Recognition*, pp. 1-12, 2017, doi: https://doi.org/10.48550/arXiv.1705.03820.
[11] V. Iglovikov and A. Shvets, "TernausNet: U-Net with VGG11 encoder pre-trained on ImageNet for image segmentation", *Computer Vision and Pattern Recognition*, 1-5, 2018, doi: https://doi.org/10.48550/arXiv.1801.05746.
[12] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation", *In: Stoyanov, D., et al. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. DLMIA ML-CDS 2018 2018*, *Lecture Notes in Computer Science, 11045. Springer, Cham.*, 2018, doi: https://doi.org/10.1007/978-3-030-00889-5_1.
[13] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi and B. H. Menze, "Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge", *Computer Vision and Pattern Recognition*, pp. 1-49, 2019, doi: https://doi.org/10.17863/CAM.38755.
[14] C. Jun, "Brain tumor dataset", *Figshare*, 2017, https://figshare.com/articles/dataset/brain_tumor_dataset/1512427.
[15] H. Mujtaba, "Introduction to ResNet or residual network", *https://www.mygreatlearning.com/blog/resnet/#sh1. Great Learning*, 2020, Access Date: 01.8.2021.
[16] A. Zhang, Z. C. Lipton, M. Li and A. J. Smola, "Dive into deep learning", *Machine Learning*, pp. 1-1185, 2020, doi: https://doi.org/10.48550/arXiv.2106.11342.
[17] Jiang, Z., Ding, C., Liu, M., Tao, D. (2020). Two-Stage Cascaded U-Net: 1st Place Solution to BraTS Challenge 2019 Segmentation Task. In: Crimi, A., Bakas, S. (eds) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2019. Lecture Notes in Computer Science, vol 11992.
[18] E. Shelhamer, J. Long and T. Darrell, "Fully convolutional networks for semantic segmentation," *in IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 640-651, 2017, doi: https://doi.org/10.1109/TPAMI.2016.2572683.
[19] L. Liu and K. Jia, "BTIS-Net: Efficient 3D U-Net for Brain Tumor Image Segmentation," in *IEEE Access*, vol. 12, pp. 133392-133405, 2024.
[20] C.Y. Lee, S. Xie, P. Gallagher, Z. Zhang and Z. Tu, "Deeply-supervised nets", *In Artificial Intelligence and Statistics*, vol. 38, pp. 562–570, 2015, doi: https://doi.org/10.48550/arXiv.1409.5185.
[21] M. Awad and R. Khanna, *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*. New York, NY, USA: Apress, 2015.
[22] L.C. Chen, G. Papandreou, I. Kokkinos, K. Murphy and A. L. Yuille, "Rethinking atrous convolution for semantic image segmentation", *Computer Vision and Pattern Recognition*, pp. 1-14, 2017, doi: https://doi.org/10.48550/arXiv.1706.05587.
[23] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications", *Computer Vision and Pattern Recognition*, pp. 1-9, 2017, doi: https://doi.org/10.48550/arXiv.1704.04861.
[24] B. Jena, S. Jain, G. K. Nayak, and S. Saxena, "Analysis of depth variation of U-NET architecture for brain tumor segmentation," *Multimedia Tools and Applications*, vol. 82, pp. 10723-10743, Sep. 2023. doi: https://doi.org/10.1007/s11042-022-13730-1
[25] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?", *IEEE*

Transactions on Medical Imaging, vol. 35, no. 5, pp. 1299-1312, 2016, doi: https://doi.org/10.1109/TMI.2016.2535302.
[26] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin and M. J. Cardoso, "Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations", *Lecture Notes in Computer Science (MICCAI)*, vol. 10553, pp. 240–248, 2017, doi: https://doi.org/10.1007/978-3-319-67558-9_28.
[27] A. Krizhevsky, I. Sutskever and E. Hinton, "ImageNet classification with deep convolutional neural networks", *Communications of the ACM,* vol. 60, pp. 84–90, 2017, doi: https://doi.org/10.1145/3065386.
[28] M. Ferguson, R. Ak, T. L. Yung-Tsun and K. H. Law, "Automatic localization of casting defects with convolutional neural networks", *IEEE International Conference on Big Data,* 2017, Boston.
[29] Jiang, Y., Zhang, Y., Lin, X., Dong, J., Cheng, T., Liang, J.: Swinbts: A method for 3d multimodal brain tumor segmentation using swin transformer. Brain Sci. 12(6), 797 (2022)
[30] Liu, X., Gao, P., Yu, T., Wang, F., Yuan, R.-Y.: Cswin-unet: Transformer unet with cross-shaped windows for medical image segmentation. Information Fusion 113, 102634 (2025)
[31] Rasool, N., Bhat, J.I., Wani, N.A., Ahmad, N., Alshara, M.: Transresunet: Revolutionizing glioma brain tumor segmentation through transformer-enhanced residual unet. IEEE Access (2024)