An attention-guided CNN framework for segmentation and grading of glioma using 3D MRI scans

Prasun Chandra Tripathi, *Student Member, IEEE,* and Soumen Bag, *Senior Member, IEEE*

Abstract—Glioma has emerged as the deadliest form of brain tumor for human beings. Timely diagnosis of these tumors is a major step towards effective oncological treatment. Magnetic Resonance Imaging (MRI) typically offers a non-invasive inspection of brain lesions. However, manual inspection of tumors from MRI scans requires a large amount of time and it is also an error-prone process. Therefore, automated diagnosis of tumors plays a crucial role in clinical management and surgical interventions of gliomas. In this study, we propose a Convolutional Neural Network (CNN)-based framework for non-invasive grading of tumors from 3D MRI scans. The proposed framework incorporates two novel CNN architectures. The first CNN architecture performs the segmentation of tumors from multimodel MRI volumes. The proposed segmentation network leverages the spatial and channel attention modules to recalibrate the feature maps across the layers. The second network utilizes the multi-task learning strategy to perform the classification based on the three glioma grading tasks which include characterization of tumor into low-grade or high-grade, identification of 1p19q, and Isocitrate Dehydrogenase (IDH) status. We have carried out several experiments to evaluate the performance of our method. Extensive experimental observations indicate that the proposed framework achieves better performance than several state-of-the-art methods. We have also executed Welch's-t test to show the statistical significance of grading results. The source code of this study is available at https://github.com/prasunc/Gliomanet.

Index Terms—Attention, Convolutional Neural Network, Glioma Grading, Magnetic Resonance Imaging, Multi-task Learning, Segmentation.

1 Introduction

Gliomas are the prevalent and lethal type of brain tumor. These tumors are classified into four grades based on aggressiveness and malignancy. The low-grade tumors (grades I-III) are typically less aggressive and exhibit a better response to the treatment. However, high-grade tumors (grade IV) have a highly aggressive nature. The high-grade tumors such as glioblastoma multiforme show poor treatment outcomes with 5 years survival for 5% patients only [1]. Early detection of this disease helps the oncologists to perform a targeted course of treatment. The detection of this disorder is typically carried out by inspecting 2D MRI slices by a radiologist. However, manual observation is typically time-consuming and vulnerable to human errors due to a large number of 2D MRI slices for each patient. Therefore, automated detection of these lesions has a crucial role in current clinical practice.

Figure 1 depicts the classification of glioma tumors based on different genetic features. As shown in this figure, every type of tumor can be classified based on the mutation in Isocitrate Dehydroge (IDH). These lesions can be characterized into two types namely, IDH mutant and IDH wild-type. The lesions which show IDH mutant typically behave better during the course of treatment in comparison to the tumors having an IDH wild-type nature. IDH mutation [2] is typically associated with 80% of low-grade and 10% of high-grade glioma tumors. The status of chromosome arm 1p/19q has a crucial significance in the treatment of low-grade lesions. The deletion of this marker is also a positive indicator for recovery [3] of the patient. Recent World Health Organization (WHO) guidelines suggest that some molecular markers along with tumor grades (low/high-grade) are also important for the treatment of tumors [4]. As illustrated in Fig. 1, molecular markers such as ATRX and TP53 are not essential for the treatment. Therefore, the classification of tumors into low-grade versus high-grade, prediction of IDH and 1p/19q status play a vital role during the medical observation provided by the oncologist.

In the current clinical practice, the status of IDH and 1p/19q is estimated from the tumor sample. The sample of the tumor is typically extracted using biopsy or tumor resection. The sample of tumor is then tested using *Fluorescence in situ hybridization (FISH)* [5]. However, the procedure of biopsy usually comprises the risk of complications [6] due to its invasive nature. Moreover, the biopsy or tumor resection may not be feasible depending upon the location and accessibility of the affected region. Thus, non-invasive prediction of tumor information is very crucial in the treatment and clinical management of glioma.

Over the last two decades, Magnetic Resonance Imaging (MRI) has become a popular tool for the inspection of different brain disorders [30] due to its ability to provide internal observations of brain tissues with high spatial and temporal resolutions. In the literature, several studies [31], [32] have been devoted to the grading of brain tumors using MRI scans. Notably, existing methods can be broadly partitioned into two groups viz., traditional and deep learnig-based methods. Traditional methods [3], [7], [8], [9], [10], [11], [12], [13] perform the characterization of tumors based on the extracted set of hand-engineered features. These methods typically utilize a machine learning classifier for the grading. For instance, researchers in [3] have utilized MR spectroscopy features extracted from multi-modal data for determining the status of 1p/19q. They have utilized random forest classifier for characterizing the tumor. Researchers have used Gray Level Co-occurrence Matrix (GLCM) for extraction of features in [7]. These features are used for

• *P. C. Tripathi is with Department of Computer Science, University of Sheffield, Sheffield, S1 4DP, United Kingdom.*
S. Bag is with the Department of Computer Science and Engineering, Indian Institute of Technology (ISM), Dhanbad, India, 826004.
E-mail: prasunchandratripathi@gmail.com

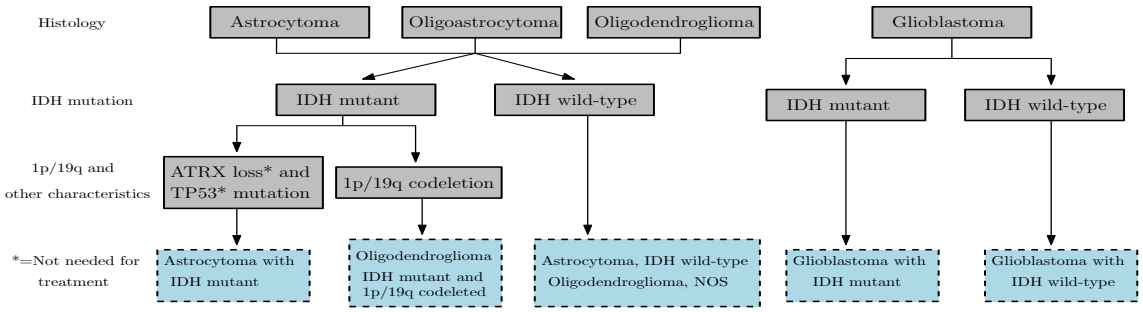


Fig. 1: The classification of brain tumors based on genetic features [4]. The tumors such as Astrocytoma, Oligoastrocytoma, Oligodendroglioma are low-grade tumors, whereas Glioblastoma is a high-grade tumor. Note that NOS represents the tumor that can not be further classified into a subgroup.

TABLE 1: The literature review of different glioma grading methods.

Sr. No.	Method	Feature set	Classifier used	Segmentation method	Type of classification	2D/3D	One-shot
1.	Fellah et al. [3]	MRI spectroscopy	Random forest	Manual	1p/19q status	2D	No
2.	Jothi et al. [7]	GLCM features	Random forest	Region growing	Low/High-grade	2D	No
3.	Hseih et al. [8]	Local and global image features	Neural network	Manual	Low/High-grade	2D	No
4.	Latif et al. [9]	Wavelet features	Random forest	Manual	Low/High-grade	3D	No
5.	Zhang et al. [10]	Shape and texture features	Random forest	Manual	IDH status	2D	No
6.	Zhou et al. [11]	Shape, histogram, and texture	Random forest	Manual	1p/19q and IDH status	2D	No
7.	Van et al. [12]	Intensity and texture fetaures	Support Vector Machine	Manual	1p/19q status	2D	No
8.	Ren et al. [13]	Radiomic features	Support Vector Machine	Manual	IDH mutation	3D	No
9.	Akkus et al. [14]	CNN features	Multi-scale CNN	Manual	1p/19q status	2D	No
10.	Yang et al. [15]	CNN features	AlexNet and GoogLeNet	Manual	Low/High-grade	2D	No
11.	Chang et al. [16]	CNN features	Residual neural network	Manual	IDH status	2D	No
12.	Ge et al. [17]	CNN features	Multi-scale CNN	Saliency-based segmentation	Low/High-grade	3D	No
12.	Sajjat et al. [18]	CNN features	VGG-19 network	Dual path CNN	Low/High-grade	2D	No
13.	Choi et al. [19]	CNN and temporal features	Recurrent Neural Network	Manual	IDH mutation	3D	No
14.	Kim et al. [20]	Texture, topological, and CNN features	Random forest	Manual	1p/19q status	2D	No
15.	Liang et al. [21]	CNN features	Dense net	Manual	IDH mutation	3D	No
16.	Bangalore et al. [22]	CNN features	Dense-CNN	Dense-Unet	IDH mutation	3D	No
17.	Zhou et al. [23]	CNN features	Recurrent Neural Network		Glioma and Meningioma	3D	Yes
18.	Sharif et al. [24]	Texture and CNN features	Inception network	Saliency-based segmentation	Low/High-grade	2D	No
19.	Ahuja et al. [25]	CNN features	VGG19 network	Superpixel-based segmentation	Low/High-grade	2D	No
20.	Tripathi and Bag [26]	CNN features	Residual network	Fuzzy-based segmentation	Low/High-grade	2D	No
21.	Neelima et al. [27]	GAN features	GAN	Unet	Low/High-grade	2D	No
22.	Ahmad et al. [28]	GAN features	GAN	GAN encoder	Low/High-grade	2D	No
23.	Cheng et al. [29]	CNN features	Multi-task CNN	Multi-task CNN	IDH status	3D	Yes

classifying the tumor into low and high-grade.

The determination of IDH status has been carried out in [33] by utilizing morphological, texture, and histogram features. Radiomics features extracted from 3D MRI scans have been used for finding IDH status in [13]. In [11], the status of 1p/19q chromosome arm has been predicted by utilizing age, histogram, shape, and texture features. The accuracy of traditional methods usually depends on the set of extracted hand-engineered features. The extraction of important features from multimodal MRI scans is a cumbersome task, which also requires the expertise of a physician. In addition to this, the choice of a suitable classifier can also impact the performance of these methods.

The deep learning-based methods [14], [15], [16], [17], [18], [19], [20], [21], [22] do not require hand-engineered features for the classification, but these methods inherently extract features based on the supervised image data. In [14], authors have introduced a novel CNN architecture for predicting 1p/19q status from MRI slices. In [15], researchers have used two pre-trained CNN architectures to perform grading of tumors into low-grade and high-grade. In [16], researchers have exploited a residual CNN for the classification of tumors from 2D MRI slices. Authors [19] have designed a short-term memory-based network for IDH classification from the multimodal MRI scans. In [21], a deep learning architecture based on dense connections has been developed for finding IDH status. A Recurrent Neural Network (RNN) has been trained on MRI features in [23] for the classification of brain lesions into three types. Researchers [25] have utilized the transfer learning on VGG19 architecture for classification of tumors. In this method, a superpixel-based segmentation approach is used. Furthermore, authors [24] have utilized Inception network to perform brain tumor classification. Recently, the fusion of four residual networks has been performed in [26] to enhance the grading performance. Generative Adversarial Network (GAN) has been utilized in [27], [28] for classification of tumors into low-grade and high-grade. A summary of various glioma grading methods has been reported in Table 1. In the literature, existing methods have been developed either for performing low/high-grade classification or for the detection of IDH and 1p/19q status. Additionally, most of these methods utilize manual or semi-automatic segmentation methods that limits the performance. In these segmentation methods, the results are often subjective in nature, which can introduce variability in the outcome [34]. The methods in [10], [18], [29] utilize automatic segmentation methods, but these methods perform grading based on a specific task. In the proposed framework, we aim to perform the grading of tumors based on three grading tasks. The main contributions of the proposed work are summarized as follows:

- We propose a novel glioma grading framework. Specifically, we develop a CNN architecture that carries out grading of brain tumors based on three tasks. The proposed classification network exploits multi-task learning for performing multiple grading tasks using a single network.

- We propose a segmentation network that segments the tumor from multimodal MRI scans. We incorporate spatial and channel attention mechanisms in the segmentation network to improve the performance.
- We perform various experiments to assess the performance of the proposed segmentation and classification networks on multimodal MRI scans. We also compare the proposed networks with existing works on the segmentation and grading of glioma.

Remaining article is divided into four sections. Section 2 briefly discusses background works on attention and multi-task learning. Section 3 explains different components of the proposed glioma grading framework. Experimental findings have been summarized in Section 4 and concluding observations have been presented in Section 5.

2 Background Works

This section provides the discussion on attention and multi-task learning in neural networks.

2.1 The Attention in CNN

The attention mechanism is utilized in neural networks to emphasize on salient information and suppress irrelevant features. This mechanism has been incorporated in different CNN architectures for solving various computer vision problems such as image classification [35], image captioning [36], semantic segmentation [37], etc. These types of CNN architectures typically utilize some learned attention vectors that are used to assign weights to different features as per their importance. In [38], authors have utilized squeeze and excitation operations to provide channel recalibration in deep neural networks. In [39], a pancreas segmentation network has been proposed . This network focuses on varying size objects by utilizing a spatial attention gating scheme. Researchers in [40] have proposed an auto-encoder-based architecture for medical image segmentation. They have utilized two parallel branches for the information flow in which one branch is specifically devoted for the attention. In [41], authors have introduced residual attention gates in their segmentation network. These attention gates help to determine discriminative features for the segmentation of medical images. Furthermore, Spatial and Channel Attention U-Net (SCAU-Net) is proposed in [42] for gland segmentation. In this method, the non-linear relationship between features is exploited by incorporating an attention mechanism in the proposed symmetrical CNN architecture. In [43], authors have incorporated channel attention in U-Net for brain tumor segmentation. This method utilizes one MRI modality for performing the segmentation which limits the performance. The attention gate has been proposed in [44] for brain lesion segmentation. The channel attention is incorporated in residual units in [45] for 2D brain tumor segmentation. Recently, researchers [46] have leveraged the attention mechanism to exploit multi-scale features in their network. They have introduced position and channel attention modules to determine useful features at different scales during the segmentation of images. The spatial and channel attention mechanisms utilized in [42], [43], [44], [45], [46] are used for 2D segmentation of the medical images. In the proposed method, we use spatial and channel attention mechanisms for obtaining segmentation of 3D tumor region from multimodal MRI scans. We have extended the channel attention mechanism in [42] to its 3D form to incorporate it into our networks. For spatial attention, we have utilized a projection tensor which defines the importance of each voxel in a patch. The projection tensor is generated by squeezing the dimensions of the input using a convolution kernel.

2.2 Multi-task Learning

Multi-task learning helps to utilize the knowledge of one task to perform learning of interrelated tasks. This form of learning can be utilized in a deep neural network using hard and soft parameter sharing [47]. In hard parameter sharing, the hidden layers are shared between the tasks, whereas task-specific layers are different for each task. In soft parameter sharing, each task has its network and the distance between parameters is optimized using a regularization approach. The multi-task learning has also been applied in some works for medical image processing. For example, researchers in [48] have developed a multi-task network which is used for determining the malignancy of lung nodules from CT images. In [49], authors have introduced multi-task CNN for cell detection in colon cancer. Researchers in [50] have developed a multi-task CNN architecture for the segmentation of brain tumors. In this work, the segmentation outcomes of two decoders are combined by utilizing multi-task learning. The segmentation and prediction of IDH status is performed in [51] using multi-task learning. Recently, authors [29] have used a CNN-Transformer in multi-task learning-based method. In this work, the tasks of segmentation and IDH mutation are performed in a single pass.

3 The Proposed Methodology

The schematic overview of the proposed framework has been depicted in Fig. 2. The proposed framework for tumor grading consists of two steps. In the first step, the glioma is segmented from the multimodel MRI volumes. We have proposed an encoder-decoder-based architecture for the segmentation of tumors. The proposed segmentation network also incorporates spatial and channel attention to recalibrate feature maps in different convolution layers. In the second step, the segmented 3D tumor region is classified using a classification network. The classification network utilizes multi-task learning to simultaneously perform grading based on the three tasks. The first task performs grading based on the low-grade and high-grade, whereas the second and third tasks characterize tumors based on the mutations in IDH and 1p/19q status. In this section, we explain the architectures, basic building blocks, and other salient features of the proposed segmentation and classification networks.

3.1 Segmentation Network

The glioma tumor is usually segmented into four intra-tumorous regions viz., edema, necrotic, non-enhancing region, and enhancing region [52]. These tumor regions are typically mapped into three tumor classes, which are essential from the treatment and surgical perspective [53]. These classes are defined as Whole Tumor (WT), Tumor Core (TC), and Enhancing Tumor (ET). WT class comprises all tumor regions, TC class includes all tumor regions except edema, and ET class specifically contains enhancing region. In this work, we have used an encoder-decoder-based architecture for the segmentation of glioma regions. Figure 3 illustrates the architecture of the backbone CNN used in the proposed method. This segmentation network takes four MRI sequences (T1, T1C, T2, and FLAIR) as input and produces a segmentation map corresponding to five regions i.e, edema, necrotic, non-enhancing region, enhancing region, and background. We have utilized different colors to represent different

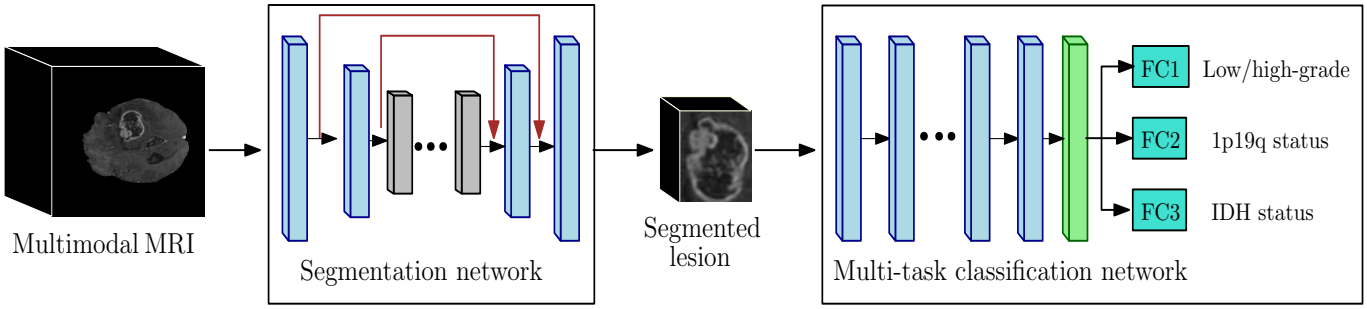


Fig. 2: The proposed framework for the grading of the tumors. FC1, FC2, and FC3 represent three fully-connected layers used in the classification network.

basic building blocks. The segmentation network leverages 3D convolution filters for the feature extraction across the layers.

The proposed segmentation network comprises three main components namely, encoder, transition, and decoder block. The encoder block downsamples the input tensor into different resolutions. This block contains three downsampling layers. Each downsampling layer reduces the size of the input tensor by a factor of 2. The encoder block also consists of some residual blocks. Every residual block contains three convolution layers. These blocks help to provide residual learning locally in the segmentation network using short-range skip connections. The transition block helps to extract deep feature representations from the encoded features. This block specifically contains a residual block that attempts to extract latent features from the reduced representations. The decoder block performs up-sampling of the input tensor to regain the resolution. It uses three up-sampling layers. The segmentation network also incorporates three long-range skip connections. These skip connections are used for preserving shallow or low-level image features. The segmentation network uses all convolutions of kernel size $3 \times 3 \times 3$ except the final kernel, which is of size $1 \times 1 \times 1$. We have extracted MRI patches of size $64 \times 64 \times 32 \times 4$ voxels from the input dataset for the training. Here, the starting three numbers show the dimension of the input volume, whereas the last number represents the four MRI channels (i.e., T1, T1C, T2, and FLAIR). We have combined the multi-class soft dice function with the cross-entropy function to form the loss function of our segmentation network. The brain tumor segmentation dataset contains only two percent of tumor tissues in the maximum number of cases [54]. Therefore, this dataset is imbalanced. If we specifically use the dice function as a loss function, then the segmentation network suffers from convergence issue in the training. To solve this issue, we have combined the cross-entropy function with the dice function, which improves the convergence of the training. The loss function can be expressed as follows:

$$loss = -\frac{1}{n_p}\sum_{i=1}^{n_p}\sum_{j=1}^{\kappa} G_i^j \log O_i^j - \frac{2}{\kappa}\sum_{j=1}^{\kappa}\frac{\sum_{i=1}^{n_p} O_i^j G_i^j}{\sum_{i=1}^{n_p} O_i^j + \sum_{i=1}^{n_p} G_i^j}, \quad (1)$$

where O denotes the predicted probability of the class, G shows *one hot vector* encoding of the ground truth, n_p represents number of voxels in a patch, and κ shows number of tumor classes. In Eq. 1, the first term represents the cross entropy loss and the second term denotes the multi-class dice loss.

3.2 Spatial and Channel Attention Mechanisms

The attention mechanism is utilized in machine learning to select the important information and ignore unnecessary things [55]. In a CNN, the attention mechanism can be applied using two ways viz., channel attention and spatial attention. A CNN produces different kinds of feature maps at each convolution layer. Each feature map or channel contributes equally to the prediction process. The channel attention mechanism is incorporated in a CNN architecture to improve the performance by providing unequal importance to different feature maps across the layers. Notably, the channels which are more relevant for the prediction get high importance in comparison to the remaining ones. On the other hand, the spatial attention mechanism is used for enhancing the importance of salient locations in the feature maps or channels.

The proposed segmentation network takes an MRI I as input and produces a segmentation map J using a mapping function $F : I \rightarrow J$. This function consists of a sequence of composite functions corresponding to the layers in encoder and decoder blocks. Each layer mainly performs convolution, pooling, and deconvolution operations. Suppose, a feature map $Y \in \mathbb{R}^{M \times N \times H \times C^{'}}$ passes through a layer that applies a function $F_r(.)$ and produces an output feature map $X \in \mathbb{R}^{M \times N \times H \times C}$. Here, $M \times N \times H$ denotes the dimensions of input MRI, $C^{'}$ represents the number of channels in the input feature map, and C shows the number of channels in the output feature map.

The channel attention utilizes squeeze and excitation operations [38] to get the channel descriptor S. This descriptor is specifically used for scaling up the prominent channels. Let us consider the input tensor $X = [x_1, x_2, ..., x_C]$ as a combination of feature maps, where $x_i \in \mathbb{R}^{M \times N \times H}$. The descriptor S is estimated by applying the function $F_{SE}(.)$ on the input feature maps X (Fig. 4a). The function F_{SE} is a composite function, which is computed as $F_{SE} = F_{EX}(F_{SQ}(X))$. To this end, we firstly perform squeeze operation on the input tensor X as follows:

$$p_l = F_{SQ}(x_l) = \frac{1}{MNH}\sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{k=1}^{H} x_l(i,j,k), \quad (2)$$

where the vector $P = [p_1, p_2, ..., p_l, ..., p_C]$ embeds the global information of each channel. Furthermore, the global information obtained in vector P is passed through two fully connected layers. These layers attempt to encode latent channel-wise dependencies in the attention. The fully connected layers transform vector P into S using the following equation:

$$S = F_{EX}(P) = W_1(\delta(W_2 P), \quad (3)$$

where W_1 and W_2 represent the weights of the first and second fully connected layers and $\delta(.)$ denotes ReLU (Rectified Linear Unit) function. The output of the last fully connected is passed through the sigmoid function $\sigma(.)$ that brings the output in the range $[0, 1]$. Therefore, we can express the output of channel attention $\hat{X}_{CA}$ using the following equation:

$$\hat{X}_{CA} = [\sigma(s_1)x_1, \sigma(s_2)x_2, ..., \sigma(s_C)x_C] \quad (4)$$

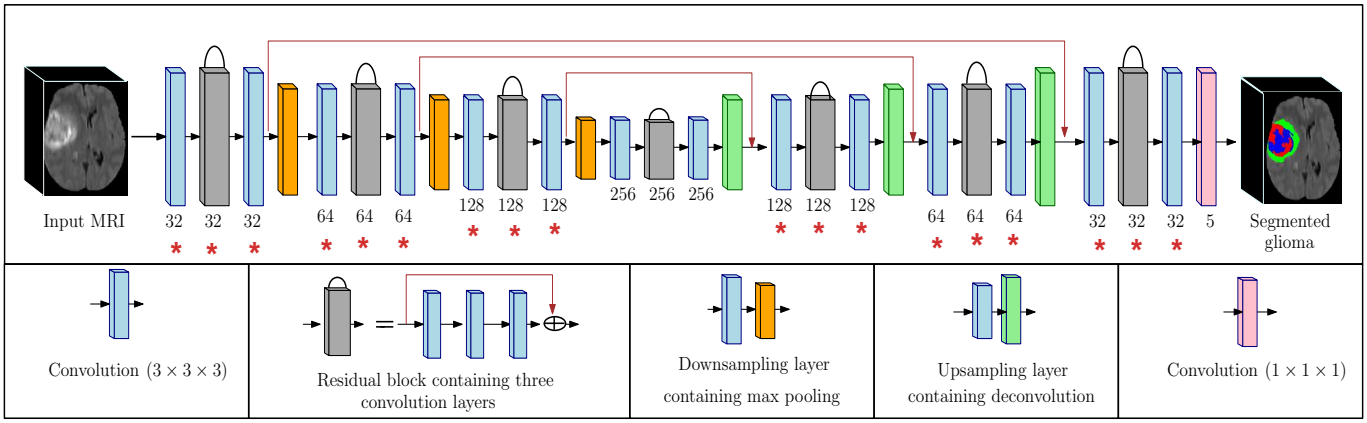
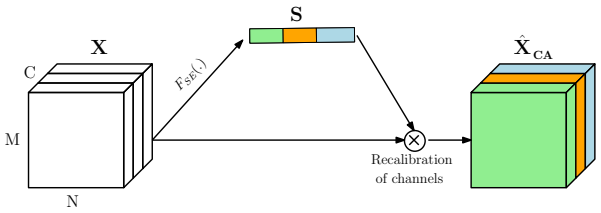
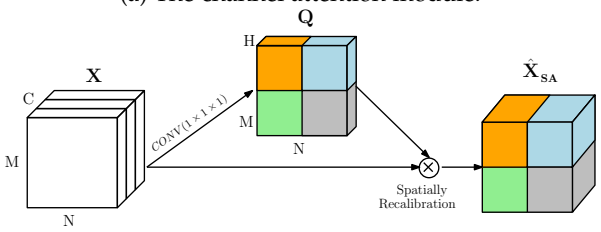


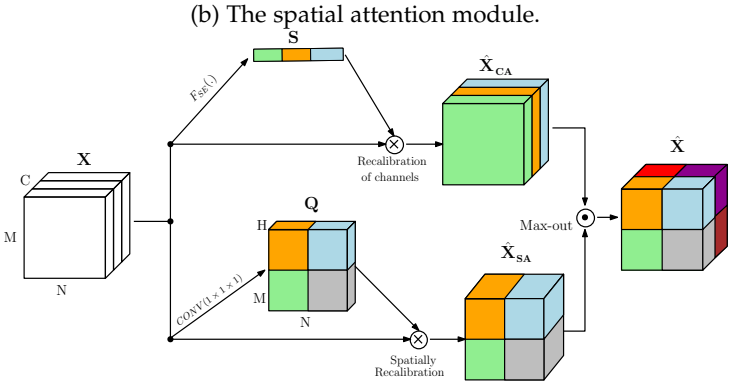
Fig. 3: The architecture of the segmentation network (Segnet). The upper part of the figure depicts the backbone segmentation network, whereas the lower part of the figure shows five basic building blocks incorporated in the network. The number of kernels in each layer has been represented with a number below of each building block. The symbol (*) below each layer represents that the layer uses attention mechanism.



(a) The channel attention module.



(b) The spatial attention module.



(c) The fusion of two types of attention modules.

Fig. 4: The architectures of attention modules incorporated in the proposed method. The input feature map X gets recalibration in each module.

In Eq. 4, the activation $\sigma(s_i)$ shows the importance of i^{th} channel. The network determines the optimum values of these activations during the training phase. The trained model performs the segmentation by suppressing the less important channels and scaling up the crucial ones.

The block diagram of the spatial attention module is illustrated in Fig. 4b. This form of attention is utilized to recalibrate or excite the specific voxel locations in the feature map for achieving fine-grained segmentation of the image. We have utilized the information from each channel of the input tensor X to estimate the importance of each spatial location. Suppose, we consider the input tensor X in the form, $X = [x^{1,1,1}, x^{1,1,2}, ..., x^{i,j,k}, ..., x^{M,N,H}]$ where $x^{i,j,k} \in \mathbb{R}^{1\times1\times1\times C}$. A projection tensor $Q \in \mathbb{R}^{M\times N\times H}$ is generated from the input feature map X by applying a convolution operator, i.e. $Q = W \star X$, where W denotes the weights of the convolution kernel and the convolution operator is represented using $\star$ symbol. The value of each voxel location $q_{i,j,k}$ of the projection tensor Q comprises the combination of each voxel from the input. We have used the sigmoid activation function $\sigma(.)$ to convert all values of the projection tensor into the range $[0, 1]$. Therefore, the output feature map $\hat{X}_{SA}$ can be represented as follows:

$$\hat{X}_{SA} = [\sigma(q_{1,1,1})x^{1,1,1}, ..., \sigma(q_{i,j,k})x^{i,j,k}, ..., \sigma(q_{M,N,H})x^{M,N,H}] \quad (5)$$

In Eq. 5, the activation $\sigma(q_{i,j,k})$ represents the importance of a location (i, j, k) of the feature map. The optimum values of these activations are estimated during the training of the network. The trained model performs the segmentation by focusing more on the salient locations of the feature maps while ignoring the useless ones.

The proposed segmentation network utilizes the combination of channel and spatial attention modules. Figure 4c depicts the fusion of the two modules. We have used max-out aggregation strategy to fuse the output of two attention modules. The output of fusion $\hat{X}$ can be represented as follows:

$$\hat{X}(i, j, k, c) = max(\hat{X}_{CA}(i, j, k, c), \hat{X}_{SA}(i, j, k, c)) \quad (6)$$

This type of fusion approach provides location-wise aggregation of two modules which helps to get selective attention in the segmentation. We can also apply the other aggregation approaches such as addition, concatenation, etc. The addition approach provides equal importance to both modules at each location, which limits the performance of the tumor segmentation. On the other hand, the concatenation of the outputs of two modules increases the number of channels, which inevitably puts a computational load on the model.

The feature maps generated from convolution layers are scaled based on the attention mechanism. Therefore, we embed the proposed attention module (Fig. 4c) in different convolution layers. The segmentation network contains three components which include encoder, transition, and decoder. The attention module can be integrated in different combinations of these components. We have utilized the attention mechanism in encoder and decoder components as this combination provides

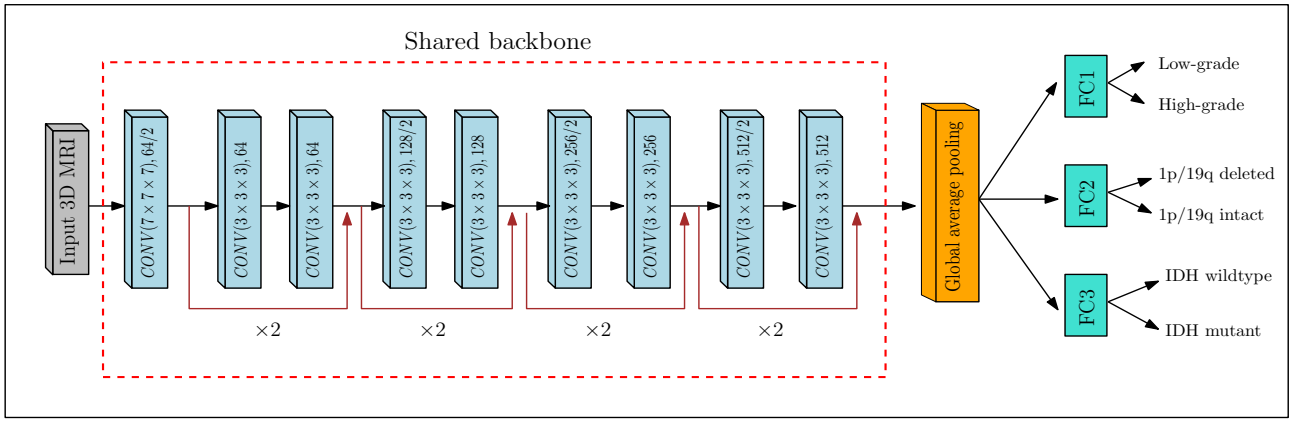


Fig. 5: The architecture of the proposed classification network. FC1, FC2, and FC3 represent fully-connected layers used for the multi-task learning. The symbol (×2) denotes that the block is repeated for two times.

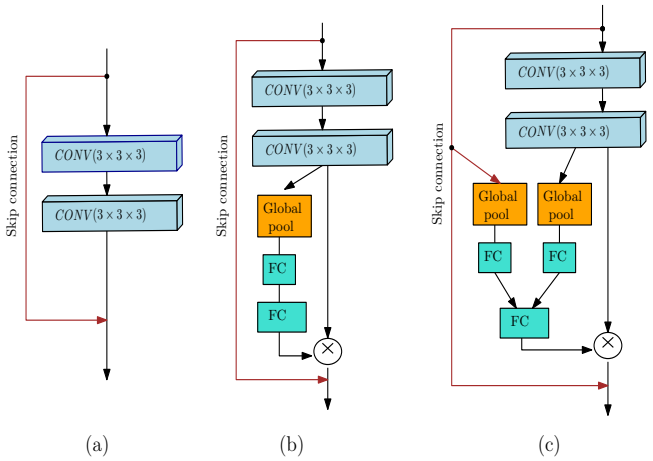


Fig. 6: Incorporating channel-wise attention in the residual block: (a) residual block, (b) Single Path Attention (SPA) block, (c) Dual Path Attention (DPA) block.

better performance than other combinations (please see Table 4). In Fig. 3, a symbol (*) below each layer depicts that the layer incorporates attention.

3.3 Multi-task Classification Network for Glioma Grading

We have designed a multi-task learning-based CNN architecture that classifies the segmented 3D glioma region. The multi-task learning performs different interrelated tasks simultaneously using a single network [56]. In this type of learning, a shared representation is utilized for the training. The tasks usually share domain-specific information among them to boost the performance. In this work, we have utilized a supervised form of multi-task learning. In this setting, every task contains a labeled dataset for the training. Suppose a task T_i contains a training dataset D_i, which has n_i training instances. We can represent the dataset D_i as $\{u_j^i, v_j^i\}_{j=1}^{n_i}$, where $u_j^i \in \mathbb{R}^{d_i}$ is j^{th} instance in task T_i and v_j^i denotes its class label. The training data matrix for task T_i has been represented using U^i, i.e., $U^i = (u_1^i, u_2^i, ..., u_{n_i}^i)$. If different tasks are defined on the same feature space, then $d_i = d_j$ for any $i \neq j$. This setting is typically known as homogeneous multi-task learning. Notably, the problem of glioma grading comprises three interrelated tasks which share the same feature space. So, we have utilized homogeneous multi-task learning in the proposed work.

The architecture of the proposed classification network is depicted in Fig. 5. The classification network contains a shared backbone and some task-specific layers. The shared backbone is utilized for extracting prominent features using different convolution layers. The first layer of the classification network is a convolution layer that consists of 64 kernels of size $7 \times 7 \times 7$. This layer also uses downsampling with the stride of 2. In order to achieve a smooth flow of gradients, the classification network utilizes residual learning using a set of residual blocks. Each residual block contains two convolution layers of size $3 \times 3 \times 3$ each. The initial two residual blocks contain 64 kernels and these blocks do not contain any pooling operation. However, the pooling operation has been carried out in the remaining residual blocks. The number of kernels is doubled after every two blocks. Therefore, the last residual block comprises 512 filters in each convolution layer.

The feature maps generated from the shared backbone are fed into a global average pooling layer. This layer specifically helps to deal with different sizes of input MRI. Therefore, we do not need to resize the segmented MRI region to feed into this network. The output of the global average pooling is passed through three fully connected layers. These layers are task-specific layers. The first fully connected layer (FC1) classifies the image based on the low-grade or high-grade tumor. The second fully connected layer (FC2) characterizes the tumor based on the status of 1p/19q, whereas the third fully connected layer (FC3) classifies the tumor based on IDH status. The training of the network is carried out by optimizing the cross-entropy function for each task. The parameters of the shared backbone are utilized by each task to perform the prediction. In this work, every task is a binary classification task. Thus, the cross-entropy loss can be defined as follows:

$$loss(\boldsymbol{v}, \hat{\boldsymbol{v}}) = -\sum_{i=1}^{n_b} \Big(v_i^j \log \hat{v}_i^j + (1 - v_i^j) \log(1 - \hat{v}_i^j) \Big), \quad (7)$$

where v_i^j is a class label for i^{th} instance, $\hat{v}_i^j$ is the predicted output of *softmax* function for i^{th} training instance of task T_j, and n_b denotes number of samples in a batch of the training. The multi-task learning scheme not only provides concurrent execution of three tasks, but it also helps to deal with some missing labels for a task during the training. The MRI data for IDH status contains several missing class labels. The proposed multi-task learning-based network allows learning important features exhaustively from the interrelated tasks to handle missing class labels for a specific task.

The proposed network incorporates attention mechanism in all residual blocks. Figure 6 illustrates three forms of residual blocks. The first block is a simple residual block that does not contain any attention module. Figure 6(b) shows Single Path Attention (SPA) block. In this block, the attention has been introduced after the second convolution layer. Due to this, the attention is specifically performed on the output of the second convolution layer, and the information from the identity or skip connection can not be utilized for the attention mechanism. To deal with this, we have also incorporated the information from skip connection into the attention mechanism. Figure 6(c) depicts Dual Path Attention (DPA) block that utilizes information from two paths for the attention mechanism. We have used a pooling and fully connected layer to handle the feature maps extracted from the skip connection. A fully connected layer has been used to combine the information extracted from the two paths. DPA block receives attention from different paths, which helps in the comprehensive selection of features, and thereby it boosts the performance.

4 Experimental Results and Analysis

We have performed several experiments to test the performance of the proposed framework. In this section, we discuss description of dataset, training of the networks, evaluation metrics, and result analysis for the segmentation and classification.

4.1 Description of Dataset

We have utilized MRI images available from open-source datasets to perform various experiments in this work. The proposed segmentation network has been trained with BRATS 19 dataset [57], [58], [59]. This dataset provides multimodal MRI data that includes T1, T2, T1C, and FLAIR modalities. The images have been segmented by expert radiologists to provide ground truth segmentation maps for different regions of glioma. For each case, four raters have segmented MRI images and the ground truth segmentation has been created based on a majority voting algorithm. The glioma regions are segmented into four subregions which include edema, non-enhancing lesion, enhancing lesion, and necrosis. The dataset provides 335 cases for different glioma patients. All scans in this dataset have resolution of $244 \times 155 \times 155$. To test the effectiveness of our classification network, we have also collected scans from The Cancer Imaging Archive (TCIA) [1] datasets. The datasets which provide glioma grading in the form of low-grade/high-grade, mutation of IDH, and 1p/19q status have been included in this study. We have used three datasets from TCIA which include TCGA-GBM, TCGA-LGG, and TCGA-1p19qdeletion. TCGA-GBM provides pre-operative MRI scans of glioblastoma multiforme that is a high-grade tumor. This dataset provides T1, T2, T1C, and FLAIR modalities. TCGA-LGG and TCGA-1p19qdeletion datasets contain low-grade tumor images. These datasets comprise pre-operative scans of grade II and III glioma of different patients. TCGA-1p19qdeletion dataset contains MRI scans of T2 and T1C modalities only.

We have acquired total MRI data of 617 patients. The cases that contain very few MRI slices for T1C, T2 or FLAIR modalities have been excluded from the result analysis. We have taken 158 cases from TCGA-GBM dataset. The low-grade scans of 257 cases have been taken from TCGA-LGG and TCGA-1p19qdeletion datasets. We have collected MRI scans for 119 cases from TCGA-LGG dataset, whereas the images for 138 patients are obtained from TCGA-1p19qdeletion dataset. The scans in BRATS 19 dataset contain 202 cases of low-grade and high-grade tumors. It should be noted that out of 335 cases of BRATS 19 dataset 133 cases are common in TCIA database. In the collected scans, the class labels are available for all patients for glioma classification into low-grade versus high-grade. There are 331 low-grade cases and 286 high-grade cases. For IDH status, we have 377 cases in which 166 patients belong to IDH-wild type and 211 cases are from the IDH-mutant group. The status for 1p19q is available for 275 low-grade tumor cases in which 129 patients have deleted status and 146 patients have intact status. We also performed pre-processing of scans. We carried out reorientation of scans based on LPS (Left-Posterior-Superior) coordinate system. We used affine registration to co-register all scans on a T1 anatomical template. The images are skull-stripped using Brain Extraction Tool (BET) [60] and resampled to $1 \times 1 \times 1$ mmm isotropic resolution.

4.2 Training of the Networks

The proposed segmentation network has been trained on BRATS 19 dataset. We have utilized MRI data of 225 patients for the training, whereas we have used 50 cases for the validation and test set each. We have taken the batch size of 4 that fits in GPU memory without degrading the speed of the training. The training has been carried for 200 epochs. We have employed Adaptive moment estimation (Adam) [61] as a learning optimizer. The learning rate is initialized with 0.001 and it is reduced with the factor of half after every 50 epochs. Some of the glioma cases for classification data contain missing modalities. For example, 1p19deletion dataset consists of T2 and T1C modalities only. To deal with this type of problem, we have randomly removed modalities for some of the cases during the training. The values for missing modalities have been filled with zeros. In doing so, the network robustly performs prediction in cases of missing modalities. The implementation of the proposed networks has been performed in Python. We have deployed Tesla K-80 GPU for carrying out the training of CNNs.

The proposed multi-task classification network performs grading based on the segmented lesion in MRI scans. The segmented whole tumor region has been used as an ROI (Region of Interest) to feed into the classification network. We have used 10-fold cross validation to perform the result analysis of the classification network. We have trained our classification network for 160 epochs. The batch size for the training has been used as 4. We have set the learning rate as 0.0001. This rate has been decayed by a factor of two after every quarter of the learning. Adam scheme has been leveraged to perform the training in the network. To avoid the over-fitting problem, we have utilized drop-out with a probability of 0.1 in the fully connected layers. Four types of data augmentation schemes are applied during the training of the classification network. We have used random flipping on three planes, shifting the intensity levels with a factor of 0.1, random rotation with -20 to 20, and random cropping of $64 \times 64 \times 64$.

4.3 Evaluation Metrics

We have utilized seven metrics for assessing the performance of the proposed segmentation and classification networks. The performance of the proposed segmentation network is tested based on four popular segmentation metrics which include Dice Similarity Coefficient (DSC), Hausdorff distance, sensitivity, and

1. https://www.cancerimagingarchive.net/

specificity. We have tested the performance of our classification network using five well-known performance metrics namely accuracy, precision, sensitivity, specificity, and F1-score.

The DSC is a popular performance measure for image segmentation. It attempts to find the overlap between the segmented outcome and the ground-truth segmentation. It is expressed as follows:

$$DSC = \frac{2TP}{2TP + FP + FN}, \tag{8}$$

where TP represents the number of true positives, FP shows the number of false positives, and FN denotes the number of false negatives. Hausdorff distance measures the distance between the surface of ground truth and the surface of segmentation outcome. It is defined as follows:

$$Haus(E, F) = max\{\sup_{a \in A} \inf_{b \in B} d(a, b), \sup_{b \in B} \inf_{a \in A} d(a, b)\}, \tag{9}$$

where a denotes a point on the surface A of the ground truth E and b represents a point on the surface B of segmentation outcome F. The function $d(a, b)$ is used to estimate the Euclidean distance between two points a and b. The low value of Hausdorff distance indicates better segmentation outcome. This measure is sensitive to outliers. Therefore, 5% extreme values are discarded while estimating this metric. Sensitivity or recall is a crucial indicator for finding the rate at which the true positives are predicted as true positives. It is defined as follows:

$$Sensitivity = \frac{TP}{TP + FN}, \tag{10}$$

Specificity is used for finding the rate at which true negatives are predicted as true negatives. We define specificity as follows:

$$Specificity = \frac{TN}{TN + FP}, \tag{11}$$

where TN shows the number of true negatives. Accuracy is most widely used measure for judging the classification performance. It is defined as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \tag{12}$$

where TN represents the number of true negatives. The high value of accuracy is usually desirable for the classification. Precision tells the rate at which the positive class is predicted. It is defined as follows:

$$Precision = \frac{TP}{TP + FP}, \tag{13}$$

F1-score is the harmonic mean of precision and recall (sensitivity). The high score of this measure indicates that the network performs better for both classes. It is defined as follows:

$$F1 - score = \frac{2 \times precision \times recall}{precision + recall}. \tag{14}$$

4.4 Results of Segmentation

We have reported the quantitative segmentation results of glioma in Table 2. The results have been reported for the segmentation of three tumor sub-regions viz., Enhancing Tumor (ET), Whole Tumor (WT), and Tumor Core (TC). In this table, we have compared the results for four CNN architectures. Our backbone segmentation architecture (Fig. 3) has been represented using Segnet. After incorporating channel (Ch) and spatial (Sp) attention, we get three more CNN architectures. We can notice from this table that we get improvement in the segmentation results when we apply the attention module in the network. The performance of the network with spatial attention (Segnet+Sp) is better than the network with channel attention (Segnet+Ch). The network also achieves improvements in the results when we apply both types of attention modules simultaneously. Therefore, we can say that the attention mechanism helps to improve the performance of tumor segmentation because it selects the crucial information from the feature maps which is relevant for this task.

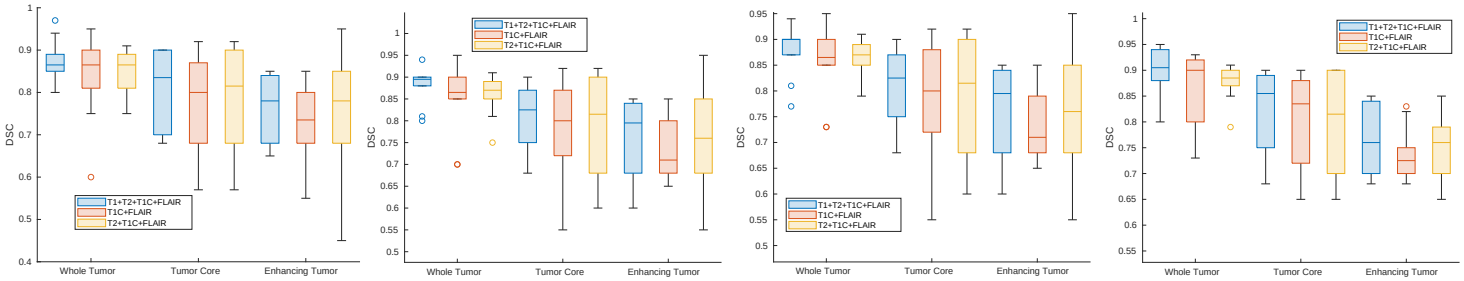
TABLE 2: The glioma segmentation results using different CNN architectures on BRATS 19 validation data. Segnet represents the proposed segmentation network without incorporating any form of attention. Sp shows spatial attention and Ch represents channel attention.

Metric	Class label	Method			
		Segnet	Segnet+Ch	Segnet+Sp	Segnet+Ch+Sp
DSC	ET	0.7612	0.7693	0.7701	**0.7712**
	WT	0.8721	0.8802	0.8829	**0.9002**
	TC	0.8090	0.8098	0.8102	**0.8230**
Sensitivity	ET	0.7662	0.7671	0.7720	**0.7870**
	WT	0.8832	0.8869	0.8871	**0.8955**
	TC	0.7625	0.7693	0.7700	**0.7710**
Specificity	ET	0.9967	0.9969	0.9972	**0.9979**
	WT	0.9956	0.9958	0.9959	**0.9962**
	TC	0.9962	0.9962	0.9965	**0.9971**
Hausdorff	ET	3.49	3.27	3.08	**2.97**
	WT	4.39	4.17	4.12	**3.73**
	TC	7.18	7.02	6.86	**5.82**

We have also performed an ablation experiment to test the performance of the network when we utilize different sets of MRI modalities in the training. We have taken three sets of MRI modalities which include T1C+FLAIR, T2+T1C+FLAIR, and T1+T2+T1C+FLAIR. The modalities T1C and FLAIR are very important for the segmentation because T1C attempts to capture enhancing region and FLAIR captures the edema region of the tumor. Due to this, these two modalities are included in every set. We have depicted the comparison of results for different sets in Fig. 7. The box plots in this figure provide the performance of different sets for segmenting three tumor regions. We can notice from these graphs that the average DSC score for the set T1+T2+T1C+FLAIR is superior among the three sets for different networks. Therefore, the proposed segmentation network provides the best performance when we utilize all MRI modalities for the training.

The visual outcomes of the glioma segmentation have been presented in Fig. 8. In this figure, we have compared the results of four segmentation networks which include Segnet, Segnet+Ch, Segnet+Sp, and Segnet+Ch+Sp. The segmented tumor regions have been depicted using three different colors. Green color has been utilized to represent the edema region, whereas Red and Blue colors have been used for representing enhancing and necrotic regions. We have represented the segmentation outcomes in a column-wise fashion. The first column represents the input MRI scan and the second column shows the ground segmentation of the tumor. The third to sixth columns represent the outcomes of four CNN networks viz., Segnet, Segnet+Ch, Segnet+Sp, and Segnet+Ch+Sp in the respective order. We have shown the enlarged regions in the second and fourth rows.

We can observe from the center of tumor regions that some of the pixels belonging to the necrotic region are incorrectly marked as enhancing region in the outcome of Segnet (Column 3) and Segnet+Ch network (Column 4). However, these pixels are correctly classified in the outcomes of Segnet+Sp and Segnet+Ch+Sp networks (Columns 5 and 6). Therefore, the spatial attention helps to improve the result of enhancing



(a) Segnet. (b) Segnet+Ch. (c) Segnet+Sp. (d) Segnet+Ch+Sp.

Fig. 7: The performance of the proposed segmentation networks on BRATS 19 validation dataset for different set of MRI modalities.

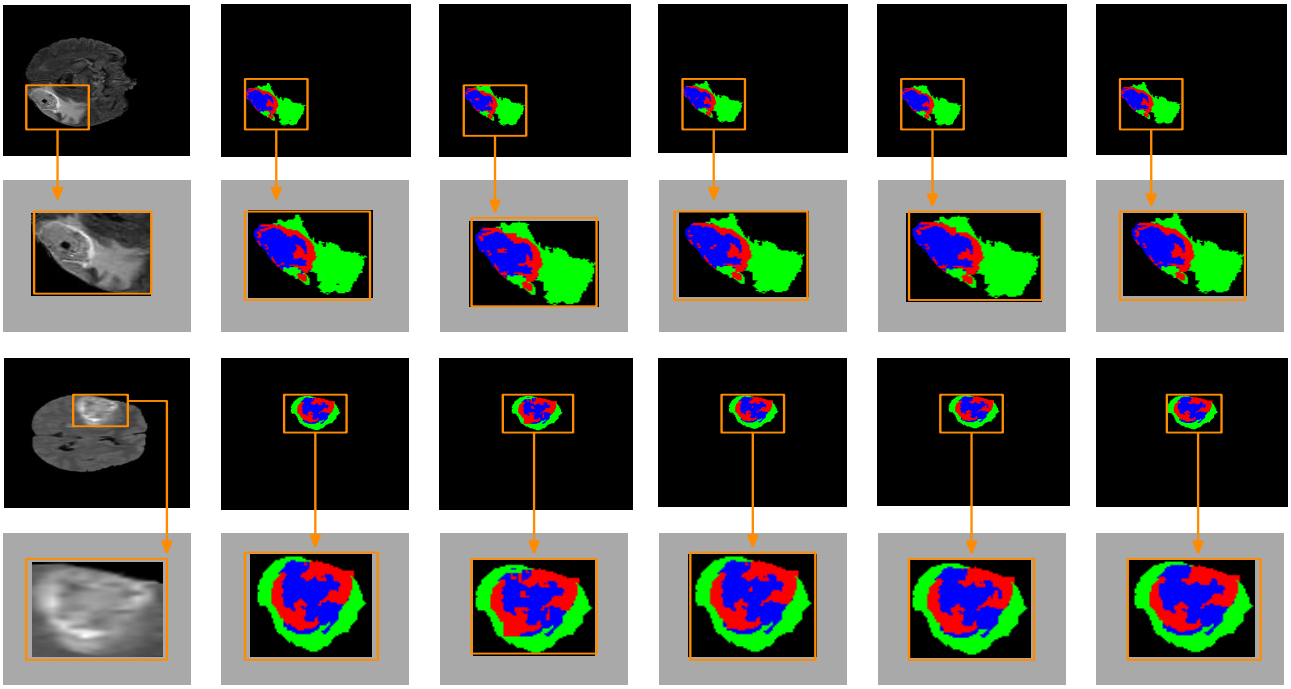


Fig. 8: The visual segmentation outcomes on two MRI scans. First column: Input MRI scans, Second column: The ground truth segmentation, Third column: Output of Segnet, Fourth column: Output of Segnet+Ch, Fifth column: Output of Segnet+Sp, Sixth column: Output of Segnet+Ch+Sp model. The second and fourth row show enlarged tumor regions.

region. If we further compare the outcomes of different CNNs then we may notice that the result of Segnet+Ch+Sp network is very promising which preserves more details of intra-tumorous regions. Interestingly, the outcome of Segnet+Ch+Sp network is very similar to the ground-truth segmentation. It can be concluded with these results that incorporation of the channel and spatial attention in the proposed segmentation network (Segnet) helps to produce fine-grained segmentation of the glioma region. Furthermore, we have compared the performance of different fusion schemes for attention modules. Table 3 reports the results of four types of fusion schemes. It is evident from this table that max-out fusion scheme outperforms other fusion schemes for tumor segmentation. Due to this, we have utilized max-out fusion scheme for performing various experiments in this work. An experiment is also performed to find the optimal placement of the proposed attention block (Fig. 4c) in the segmentation network. We have explored five type of combinations of placement of attention block. The results of this experiment are reported in Table 4. We can observe from this table that we get better performance when we place the attention block in encoder and decoder.

TABLE 3: The comparison of different fusion schemes based on DSC score and Hausdorff distance.

Fusion method	DSC			Hausdorff		
	ET	WT	TC	ET	WT	TC
Addition	0.7602	0.8767	0.8169	3.35	4.13	6.41
Multiplication	0.7654	0.8889	0.8204	3.32	3.92	6.13
Concatenation	0.7705	0.8995	0.8223	3.08	3.81	5.91
Max-out	**0.7712**	**0.9002**	**0.8230**	**2.97**	**3.73**	**5.82**

TABLE 4: The comparison for different positions of attention block in the segmentation network.

Attention	DSC			Hausdorff		
	ET	WT	TC	ET	WT	TC
Only encoder	0.7622	0.8792	0.8174	3.30	4.30	6.05
Only decoder	0.7619	0.8788	0.81804	3.49	4.91	5.94
Only transition	0.7620	0.8752	0.8149	3.50	4.21	6.13
Encoder and decoder	**0.7712**	**0.9002**	**0.8230**	**2.97**	**3.73**	**5.82**
Encoder, transition, and decoder	0.7702	0.8931	0.8220	3.03	3.95	5.92

TABLE 5: The comparative analysis of seven glioma segmentation methods on BRATS 19 validation data. Best performing scores are highlighted as bold.

Method	DSC			Hausdorff		
	ET	WT	TC	ET	WT	TC
Pereira et al. [62]	0.76	0.89	0.78	3.14	4.28	6.98
Ding et al. [63]	0.67	0.85	0.73	6.54	5.12	6.24
Akil et al. [64]	0.73	0.87	0.73	5.05	4.81	6.48
Kamnitsas et al. [65]	0.63	0.86	0.68	5.19	4.87	6.91
Isensee et al. [66]	**0.80**	0.89	**0.82**	**2.62**	4.07	5.93
Henry et al. [67]	0.77	0.88	**0.82**	3.02	4.13	5.97
Proposed method	0.77	**0.90**	**0.82**	2.97	**3.73**	**5.82**

TABLE 6: The performance of the proposed network for three glioma grading tasks using three types of residual blocks.

Metric	Task	Method		
		Residual block	SPA block	DPA block
Accuracy	Low/High-grade	91.00	94.50	**95.86**
	IDH status	87.94	90.97	**91.96**
	1p19q status	81.20	86.03	**87.88**
Precision	Low/High-grade	92.10	93.51	**94.86**
	IDH status	89.20	90.58	**93.54**
	1p19q status	84.25	88.12	**89.23**
Specificity	Low/High-grade	93.31	94.53	**96.81**
	IDH status	90.51	92.56	**93.48**
	1p19q status	83.56	85.59	**91.57**
Recall	Low/High-grade	92.13	93.56	**94.82**
	IDH status	88.12	90.18	**90.98**
	1p19q status	85.61	86.87	**89.57**
F1-score	Low/High-grade	92.13	93.53	**94.84**
	IDH status	88.65	90.41	**92.24**
	1p19q status	84.92	87.49	**89.39**

We have compared our segmentation network against six recent glioma segmentation methods [62], [63], [64], [65], [66], [67]. These methods have been compared on BRATS 19 dataset. We have reproduced the results for each comparing method on BRATS 19 dataset to provide unbiased comparison. Table 5 reports the performance of different methods in terms of DSC score and Hausdorff distance. The table summarizes the results for three tumor classes namely, ET, WT, and TC. It can be observed from the reported results that the proposed method provides highly competitive performance. The proposed method produces the best DSC score for WT segmentation. Isensee et al.'s [66] method provides the best DSC score for ET segmentation. Furthermore, we can observe from the table that the proposed method provides minimum hausdorff distance for WT and TC segmentation.

4.5 Results of Classification

We have reported the performance of the classification in Table 6. This table incorporates the performance of three grading tasks for five performance metrics. We have also reported the performance of the classification network when we utilize different types of residual blocks (see Fig. 6) in the proposed classification architecture (see Fig. 5). We can observe from this table that the proposed network achieves promising results for glioma classification. If we compare the result of three types of blocks then we can notice that the attention-based blocks (SPA and DPA) perform better in comparison to the simple residual block. However, the dual path attention or DPA block outperforms among three blocks. The single attention path-based block does not fully exploit the attention mechanism. On the other hand, dual path-based attention helps to exhaustively select the important features for the classification. Due to this, DPA block produces superior performance for the grading of lesions.

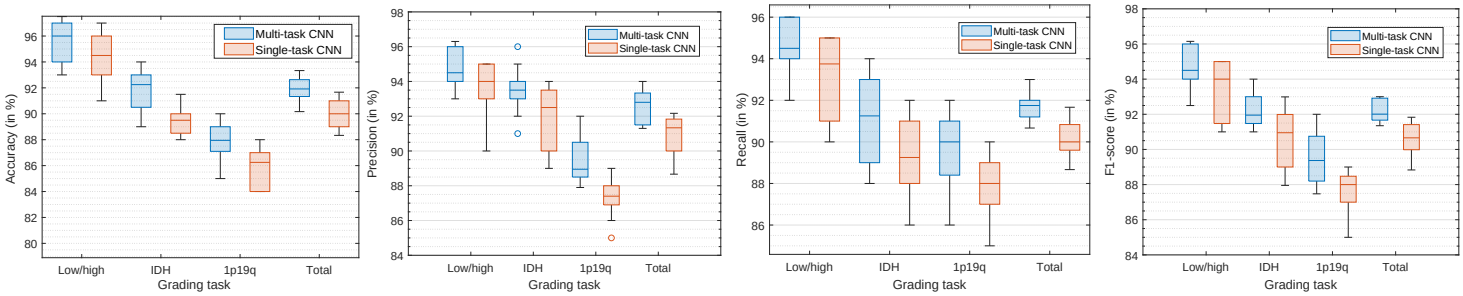
TABLE 7: The results of an ablation experiment for combining different functionality in the proposed method.

Metric	Task	Method			
		E2E-S	E2E-MT	Proposed-S	Proposed-MT
Accuracy	Low/High-grade	91.27	92.77	94.62	**95.86**
	IDH status	81.56	84.69	89.53	**91.96**
	1p19q status	75.64	79.12	85.95	**87.88**
Precision	Low/High-grade	88.79	89.52	93.65	**94.86**
	IDH status	83.16	84.61	91.95	**93.54**
	1p19q status	76.48	77.56	87.32	**89.23**
Specificity	Low/High-grade	85.78	89.52	92.91	**96.81**
	IDH status	82.92	84.16	90.23	**93.48**
	1p19q status	72.35	75.02	89.90	**91.57**
Recall	Low/High-grade	89.17	90.24	93.15	**94.82**
	IDH status	80.67	82.76	89.35	**90.98**
	1p19q status	75.18	77.88	88.03	**89.57**
F1-score	Low/High-grade	88.98	89.97	93.37	**94.84**
	IDH status	81.89	83.67	90.62	**92.24**
	1p19q status	75.82	77.72	89.00	**89.39**

We have also performed an ablation experiment to analyze the performance of the proposed classification network when we do not use multi-task learning. We have used a single FC layer in the proposed classification network (Fig. 5) to perform this experiment. We have trained three networks for each task of grading. The comparison results for this experiment are reported in Fig. 9. The figure provides the comparison in the performance based on three grading tasks. We can observe from these graphs that the performance of multi-task CNN is better than single-task CNN for three tasks of grading. Additionally, multi-task CNN produces superior performance for total classification. Therefore, we can say that multi-task learning enhances the performance.We have compared the performance of the proposed network with its 2D form. We have modified all 3D kernels to its 2D form in the proposed network. The comparison in the performance is reported in Fig. 10. It can be observed from this figure that 3D CNN produces better performance than 2D CNN for all glioma grading tasks. Therefore, the proposed 3D CNN leverages more information present in MRI scans to enhance the performance.

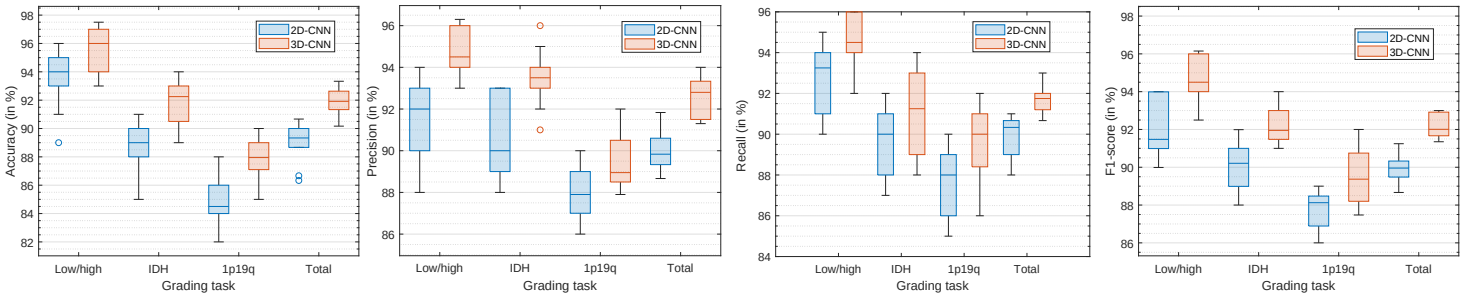
Furthermore, an ablation experiment is carried out to show the advantage of combining segmentation and classification. We have also evaluated the performance of this combination for single-task and multi-task learning. The results of this experiment have been reported in Table 7. In this table, E2E-S and E2E-MT represent networks which perform end to end classification without using the segmentation. The first network (i.e., E2E-S) uses single-task learning, whereas the second network (i.e., E2E-MT) leverages multi-task learning. We have used the segmentation in Proposed-S and Proposed-MT networks. Proposed-S network utilizes single-task learning, whereas Proposed-MT exploits multi-task learning. If we compare the results then we can see that the networks (Proposed-S and Proposed-MT) produce better performance than the networks (E2E-S and E2E-MT). Therefore, we get improved performance when we utilize segmentation of tumor region. We can also observe from these results that the performance achieved by multi-task learning networks (i.e., E2E-MT and Proposed-MT) is better than single-task learning networks (i.e., E2E-S and Proposed-S). Therefore, the utilization of multi-task learning with segmentation helps to achieve enhanced performance in the proposed method.

The performance of the proposed method has been compared with a single shot Mask RCNN architecture [68]. We have used Resnet101 as a backbone network in Mask RCNN architecture. We have utilized two popular approaches [69] to train Mask RCNN architecture. In the first approach, we have trained the entire network using the glioma dataset in an end-to-end way.



(a) Accuracy. (b) Precision. (c) Recall. (d) F1-score.

Fig. 9: The performance comparison between a single-task CNN and the proposed multi-task CNN for three glioma grading tasks.



(a) Accuracy. (b) Precision. (c) Recall. (d) F1-score.

Fig. 10: The performance comparison between 2D-CNN and 3D-CNN for three glioma grading tasks.

In the second approach, the network heads are trained first and then the entire network is trained. In this approach, the learning rate is reduced by a factor 10 after training of network heads. The network has been trained by utilizing learning rate 0.0001 with momentum 0.9. Table 8 reports the comparison results in terms of DSC and accuracy metrics. In this table, Mask RCNN 1 and Mask RCNN 2 represent the networks trained using the first and second approach respectively. It can be observed from the table that Mask RCNN 1 and Mask RCNN 2 obtain low performance in comparison to the proposed method. The glioma tumors have very complex boundaries due to their infiltrative nature. Mask RCNN architecture fails to capture the crucial tumor information and it produces coarse segmentation of the tumor. The proposed method performs separate training for the segmentation and classification of tumors that helps to learn complex tumor information.

TABLE 8: The comparison of the proposed method with single shot networks.

Mask RCNN 1		Mask RCNN 2		Proposed method	
DSC	Accuracy	DSC	Accuracy	DSC	Accuracy
0.8552	87.96	0.8618	88.37	**0.9002**	**95.86**

Finally, we have compared the performance of the proposed classification network with state-of-the-art methods for glioma grading. In Table 9, we have provided a detailed comparison of different methods for three grading tasks. The methods are compared based on the average performance achieved by them. We have used 10-fold cross validation in all methods in order to perform unbiased comparison. We can observe from the reported results that the proposed method produces better performance than existing method for every task of glioma grading. Furthermore, the statistical testing is performed to find the significance of comparison results. We have carried out Welch's t-test [70] at 5% significance level. This test compares two methods at a time to determine the results are statistically different. We have compared the results obtained by the proposed method with existing methods. In Table 10, we have reported p-values and t-values computed by this test for comparing methods. It can be noticed from the table that p-values obtained by comparing methods are less than 0.05 at 5% significance level. According to Welch's test, obtained p-values indicate that results achieved by the proposed method are statistically different and better than existing methods.

The existing CNN-based methods [14], [16], [17], [18] employ a specific network for performing a single grading task. This network is trained for same grading task to produce benchmark performance. However, the proposed classification network employs multi-task learning to perform the grading based on the three tasks. The shared representation of features is utilized to carry out grading based on the three tasks. Additionally, the proposed classification network can utilize supervised data of one task to help in the training of other tasks which contain fewer samples or some missing class labels. It has been observed that the proposed classification network produces superior performance than existing methods for all grading tasks. Due to this, the proposed network can be a favorable choice for grading of glioma in the clinical practice. The limitation of this study is that we utilize two networks for segmentation and classification. Specifically, the multi-task learning has been leveraged in the classification network. Recent work proposed by Cheng et al. [29] performs segmentation and classification in a single network by utilizing multi-task learning. In this work, spatial and global feature representations have been extracted using a

TABLE 9: The detailed comparative analysis of glioma grading for three types of tasks. Best performing scores are highlighted as bold.

Task	Method	2D/3D	Methodology Adopted	Overall Accuracy (in %)
Low-grade/High-grade	Sajjad et al. [18]	2D	CNN using VGG19	91.38
	Chenjie et al. [17]	3D	Multi-scale CNN	90.64
	Tripathi and Bag [26]	2D	Fusion of Residual networks	95.31
	Proposed method	3D	Attention-based multi-task CNN	**95.86**
IDH status	Hsieh et al. [33]	2D	Morphological, intensity and texture features	87.14
	Zhang et al. [10]	2D	Parametric intensity, texture, and shape features with random forest	89.78
	Chang et al. [16]	2D	Residual CNN	90.39
	Proposed method	3D	Attention-based multi-task CNN	**91.96**
1p/19q status	Van et al. [12]	2D	Radiomics features with support vector machine	74.80
	Akkus et al. [14]	2D	Multi-scale CNN	87.81
	Tripathi and Bag [26]	2D	Fusion of Residual networks	84.20
	Proposed method	3D	Attention-based multi-task CNN	**87.88**

TABLE 10: The result of statistical significance test.

Task	Method	t-value	p-value
Low-grade/High-grade	Sajjad et al. [18]	-9.09272	0.00012
	Chenjie et al. [17]	-8.34100	0.00013
	Tripathi and Bag [26]	-3.09204	0.00673
IDH status	Hsieh et al. [33]	-5.02459	0.00042
	Zhang et al. [10]	-6.03912	0.00091
	Chang et al. [16]	-4.05241	0.00024
1p/19q status	Van et al. [12]	-7.02123	0.00019
	Akkus et al. [14]	-8.04560	0.00175
	Tripathi and Bag [26]	-5.48124	0.00027

CNN-Transformer encoder to perform segmentation and IDH classification. However, this method does not perform glioma grading based on low-grade/high-grade and 1p/19q status. In the future, we will attempt to develop a CNN-Transformer-based architecture that can perform segmentation and grading of tumors based on three tasks of glioma grading.

5 Conclusion and Future Research

In this work, we have introduced a complete framework for the diagnosis of glioma lesions. The proposed framework provides the segmentation and classification of tumors using two CNN architectures. The attention technique is incorporated in the segmentation and classification networks to provide importance to prominent features and suppress unnecessary ones. The proposed classification network can perform different tasks simultaneously by utilizing the multi-task learning scheme. The performance of the proposed framework has been tested on a large set of multimodal MRI scans. The experimental observations reveal that the proposed glioma grading framework offers superior performance against state-of-the-art methods. The proposed method works in a supervised setting of multi-task learning. In the future, unsupervised and semi-supervised settings of multi-task learning can be used to perform grading of glioma.

Acknowledgement

We are thankful to Dr. M. K. Jha, Radiologist, Avishkar Diagnostics Dhanbad for providing clinical inputs for completing this research work. We also thank anonymous reviewers for their very good comments which helped us to improve the manuscript.

References

[1] Q. T. Ostrom, H. Gittleman, G. Truitt, A. Boscia, C. Kruchko, and J. S. Barnholtz-Sloan, "CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the united states in 2011–2015," *Neuro-oncology*, vol. 20, no. 4, pp. 1–86, 2018.

[2] H. Yan, D. W. Parsons, G. Jin, R. McLendon, B. A. Rasheed, W. Yuan, I. Kos, I. Batinic-Haberle, S. Jones, G. J. Riggins *et al.*, "IDH1 and IDH2 mutations in gliomas," *New England Journal of Medicine*, vol. 360, no. 8, pp. 765–773, 2009.

[3] S. Fellah, D. Caudal, A. M. De Paula, P. Dory-Lautrec, D. Figarella-Branger, O. Chinot, P. Metellus, P. J. Cozzone, S. Confort-Gouny, B. Ghattas *et al.*, "Multimodal mr imaging (diffusion, perfusion, and spectroscopy): is it possible to distinguish oligodendroglial tumor grade and 1p/19q codeletion in the pretherapeutic diagnosis?" *American Journal of Neuroradiology*, vol. 34, no. 7, pp. 1326–1333, 2013.

[4] D. N. Louis, A. Perry, G. Reifenberger, A. Von Deimling, D. Figarella-Branger, W. K. Cavenee, H. Ohgaki, O. D. Wiestler, P. Kleihues, and D. W. Ellison, "The 2016 world health organization classification of tumors of the central nervous system: a summary," *Acta Neuropathologica*, vol. 131, no. 6, pp. 803–820, 2016.

[5] D. Scheie, P. A. Andresen, M. Cvancarova, A. S. Bø, E. Helseth, K. Skullerud, and K. Beiske, "Fluorescence in situ hybridization (FISH) on touch preparations: a reliable method for detecting loss of heterozygosity at 1p and 19q in oligodendroglial tumors," *The American Journal of Surgical Pathology*, vol. 30, no. 7, pp. 828–837, 2006.

[6] R. J. Jackson, G. N. Fuller, D. Abi-Said, F. F. Lang, Z. L. Gokaslan, W. M. Shi, D. M. Wildrick, and R. Sawaya, "Limitations of stereotactic biopsy in the initial management of gliomas," *Neuro-oncology*, vol. 3, no. 3, pp. 193–200, 2001.

[7] G. Jothi *et al.*, "Hybrid tolerance rough set–firefly based supervised feature selection for MRI brain tumor image classification," *Applied Soft Computing*, vol. 46, pp. 639–651, 2016.

[8] K. L.-C. Hsieh, C.-M. Lo, and C.-J. Hsiao, "Computer-aided grading of gliomas based on local and global mri features," *Computer Methods and Programs in Biomedicine*, vol. 139, pp. 31–38, 2017.

[9] G. Latif, M. M. Butt, A. H. Khan, O. Butt, and D. A. Iskandar, "Multiclass brain glioma tumor classification using block-based 3D wavelet features of MR images," in *International Conference on Electrical and Electronic Engineering*, 2017, pp. 333–337.

[10] B. Zhang, K. Chang, S. Ramkissoon, S. Tanguturi, W. L. Bi, D. A. Reardon, K. L. Ligon, B. M. Alexander, P. Y. Wen, and R. Y. Huang, "Multimodal MRI features predict isocitrate dehydrogenase genotype in high-grade gliomas," *Neuro-oncology*, vol. 19, no. 1, pp. 109–117, 2017.

[11] H. Zhou, K. Chang, H. X. Bai, B. Xiao, C. Su, W. L. Bi, P. J. Zhang, J. T. Senders, M. Vallières, V. K. Kavouridis *et al.*, "Machine learning reveals multimodal MRI patterns predictive of isocitrate dehydrogenase and 1p/19q status in diffuse low-and high-grade gliomas," *Journal of Neuro-oncology*, vol. 142, no. 2, pp. 299–307, 2019.

[12] S. R. van der Voort, F. Incekara, M. M. Wijnenga, G. Kapas, M. Gardeniers, J. W. Schouten, M. P. Starmans, R. N. Tewarie, G. J. Lycklama, P. J. French *et al.*, "Predicting the 1p/19q codeletion status of presumed low-grade glioma with an externally validated machine learning algorithm," *Clinical Cancer Research*, vol. 25, no. 24, pp. 7455–7462, 2019.

[13] Y. Ren, X. Zhang, W. Rui, H. Pang, T. Qiu, J. Wang, Q. Xie, T. Jin, H. Zhang, H. Chen *et al.*, "Noninvasive prediction of idh1 mutation and atrx expression loss in low-grade gliomas using multiparametric mr radiomic features," *Journal of Magnetic Resonance Imaging*, vol. 49, no. 3, pp. 808–817, 2019.

[14] Z. Akkus, I. Ali, J. Sedlář, J. P. Agrawal, I. F. Parney, C. Giannini, and B. J. Erickson, "Predicting deletion of chromosomal arms 1p/19q in low-grade gliomas from MR images using machine intelligence," *Journal of Digital Imaging*, vol. 30, no. 4, pp. 469–476, 2017.

[15] Y. Yang, L.-F. Yan, X. Zhang, Y. Han, H.-Y. Nan, Y.-C. Hu, B. Hu, S.-L. Yan, J. Zhang, D.-L. Cheng *et al.*, "Glioma grading on conventional mr images: a deep learning study with transfer learning," *Frontiers in Neuroscience*, vol. 12, p. 804, 2018.

[16] K. Chang, H. X. Bai, H. Zhou, C. Su, W. L. Bi, E. Agbodza, V. K. Kavouridis, J. T. Senders, A. Boaro, A. Beers *et al.*, "Residual convolutional neural network for the determination of IDH status in low-and high-grade gliomas from MR imaging," *Clinical Cancer Research*, vol. 24, no. 5, pp. 1073–1081, 2018.

[17] C. Ge, Q. Qu, I. Y.-H. Gu, and A. S. Jakola, "3D multi-scale convolutional networks for glioma grading using MR images," in *International Conference on Image Processing (ICIP)*, 2018, pp. 141–145.

[18] M. Sajjad, S. Khan, K. Muhammad, W. Wu, A. Ullah, and S. W. Baik, "Multi-grade brain tumor classification using deep cnn with extensive data augmentation," *Journal of Computational Science*, vol. 30, pp. 174–182, 2019.

[19] K. S. Choi, S. H. Choi, and B. Jeong, "Prediction of IDH genotype in gliomas with dynamic susceptibility contrast perfusion MR imaging using an explainable recurrent neural network," *Neuro-oncology*, vol. 21, no. 9, pp. 1197–1209, 2019.

[20] D. Kim, N. Wang, V. Ravikumar, D. Raghuram, J. Li, A. Patel, R. E. Wendt III, G. Rao, and A. Rao, "Prediction of 1p/19q codeletion in diffuse glioma patients using pre-operative multiparametric magnetic resonance imaging," *Frontiers in Computational Neuroscience*, vol. 13, p. 52, 2019.

[21] S. Liang, R. Zhang, D. Liang, T. Song, T. Ai, C. Xia, L. Xia, and Y. Wang, "Multimodal 3d densenet for idh genotype prediction in gliomas," *Genes*, vol. 9, no. 8, p. 382, 2018.

[22] C. G. Bangalore Yogananda, B. R. Shah, M. Vejdani-Jahromi, S. S. Nalawade, G. K. Murugesan, F. F. Yu, M. C. Pinho, B. C. Wagner, B. Mickey, T. R. Patel *et al.*, "A novel fully automated mri-based deep-learning method for classification of idh mutation status in brain gliomas," *Neuro-oncology*, vol. 22, no. 3, pp. 402–411, 2020.

[23] Y. Zhou, Z. Li, H. Zhu, C. Chen, M. Gao, K. Xu, and J. Xu, "Holistic brain tumor screening and classification based on densenet and recurrent neural network," in *International MICCAI Brainlesion Workshop*. Springer, 2018, pp. 208–217.

[24] M. I. Sharif, J. P. Li, M. A. Khan, and M. A. Saleem, "Active deep neural network features selection for segmentation and recognition of brain tumors using MRI images," *Pattern Recognition Letters*, vol. 129, pp. 181–189, 2020.

[25] S. Ahuja, B. Panigrahi, and T. Gandhi, "Transfer learning based brain tumor detection and segmentation using superpixel technique," in *2020 International Conference on Contemporary Computing and Applications*, 2020, pp. 244–249.

[26] P. C. Tripathi and S. Bag, "A computer-aided grading of glioma tumor using deep residual networks fusion," *Computer Methods and Programs in Biomedicine*, vol. 215, p. 106597, 2022.

[27] G. Neelima, D. R. Chigurukota, B. Maram, and B. Girirajan, "Optimal deepmrseg based tumor segmentation with gan for brain tumor classification," *Biomedical Signal Processing and Control*, vol. 74, p. 103537, 2022.

[28] B. Ahmad, J. Sun, Q. You, V. Palade, and Z. Mao, "Brain tumor classification using a combination of variational autoencoders and generative adversarial networks," *Biomedicines*, vol. 10, no. 2, p. 223, 2022.

[29] J. Cheng, J. Liu, H. Kuang, and J. Wang, "A fully automated multimodal mri-based multi-task learning for glioma segmentation and idh genotyping," *IEEE Transactions on Medical Imaging*, 2022.

[30] A. Tiwari, S. Srivastava, and M. Pant, "Brain tumor segmentation and classification from magnetic resonance images: Review of selected methods from 2014 to 2019," *Pattern Recognition Letters*, vol. 131, pp. 244–260, 2020.

[31] G. Mohan and M. M. Subashini, "MRI based medical image analysis: Survey on brain tumor grade classification," *Biomedical Signal Processing and Control*, vol. 39, pp. 139–161, 2018.

[32] K. Muhammad, S. Khan, J. Del Ser, and V. H. C. de Albuquerque, "Deep learning for multigrade brain tumor classification in smart healthcare systems: A prospective survey," *IEEE Transactions on Neural Networks and Learning Systems*, 2020.

[33] K. L.-C. Hsieh, C.-Y. Chen, and C.-M. Lo, "Radiomic model for predicting mutations in the isocitrate dehydrogenase gene in glioblastomas," *Oncotarget*, vol. 8, no. 28, p. 45888, 2017.

[34] S. Luo, R. Li, S. Ourselin *et al.*, "A new deformable model using dynamic gradient vector flow and adaptive balloon forces," in *APRS Workshop on Digital Image Computing*, 2003.

[35] W. Cai and Z. Wei, "Remote sensing image classification based on a cross-attention mechanism and graph convolution," *IEEE Geoscience and Remote Sensing Letters*, 2020.

[36] C. Yan, Y. Tu, X. Wang, Y. Zhang, X. Hao, Y. Zhang, and Q. Dai, "Stat: Spatial-temporal attention mechanism for video captioning," *IEEE Transactions on Multimedia*, vol. 22, no. 1, pp. 229–241, 2019.

[37] Q. Zhao, J. Liu, Y. Li, and H. Zhang, "Semantic segmentation with attention mechanism for remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, 2021.

[38] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7132–7141.

[39] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz *et al.*, "Attention u-net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.

[40] C. Kaul, S. Manandhar, and N. Pears, "Focusnet: An attention-based fully convolutional network for medical image segmentation," in *IEEE International Symposium on Biomedical Imaging*, 2019, pp. 455–458.

[41] D. Lin, Y. Li, T. L. Nwe, S. Dong, and Z. M. Oo, "Refineu-net: Improved u-net with progressive global feedbacks and residual attention guided local refinement for medical image segmentation," *Pattern Recognition Letters*, vol. 138, pp. 267–275, 2020.

[42] P. Zhao, J. Zhang, W. Fang, and S. Deng, "Scau-net: Spatial-channel attention u-net for gland segmentation," *Frontiers in Bioengineering and Biotechnology*, vol. 8, p. 670, 2020.

[43] M. Noori, A. Bahri, and K. Mohammadi, "Attention-guided version of 2d unet for automatic brain tumor segmentation," in *International Conference on Computer and Knowledge Engineering*. IEEE, 2019, pp. 269–275.

[44] J. Zhang, Z. Jiang, J. Dong, Y. Hou, and B. Liu, "Attention gate resu-net for automatic mri brain tumor segmentation," *IEEE Access*, vol. 8, pp. 58 533–58 545, 2020.

[45] J. Zhang, X. Lv, H. Zhang, and B. Liu, "Aresu-net: Attention residual u-net for brain tumor segmentation," *Symmetry*, vol. 12, no. 5, p. 721, 2020.

[46] A. Sinha and J. Dolz, "Multi-scale self-guided attention for medical image segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 1, pp. 121–130, 2020.

[47] S. Ruder, "An overview of multi-task learning in deep neural networks," *arXiv preprint arXiv:1706.05098*, 2017.

[48] X. Li, Y. Kao, W. Shen, X. Li, and G. Xie, "Lung nodule malignancy prediction using multi-task convolutional neural network," in *Medical Imaging 2017: Computer-Aided Diagnosis*, vol. 10134. International Society for Optics and Photonics, 2017, p. 1013424.

[49] A. Chamanzar and Y. Nie, "Weakly supervised multi-task learning for cell detection and segmentation," in *International Symposium on Biomedical Imaging*, 2020, pp. 513–516.

[50] H. Huang, G. Yang, W. Zhang, X. Xu, W. Yang, W. Jiang, and X. Lai, "A deep multi-task learning framework for brain tumor segmentation," *Frontiers in Oncology*, p. 2095, 2021.

[51] Y. Wang, Y. Wang, C. Guo, S. Zhang, and L. Yang, "Sgpnet: a three-dimensional multitask residual framework for segmentation and idh genotype prediction of gliomas," *Computational Intelligence and Neuroscience*, vol. 2021, 2021.

[52] C. Ma, G. Luo, and K. Wang, "Concatenated and connected random forests with multiscale patch driven active contour model for automated brain tumor segmentation of MR images," *IEEE Transactions on Medical Imaging*, vol. 37, no. 8, pp. 1943–1954, 2018.

[53] T. Zhou, S. Canu, P. Vera, and S. Ruan, "Brain tumor segmentation with missing modalities via latent multi-source correlation representation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2020, pp. 533–541.

[54] D. Zhang, G. Huang, Q. Zhang, J. Han, J. Han, Y. Wang, and Y. Yu, "Exploring task structure for brain tumor segmentation from multi-modality MR images," *IEEE Transactions on Image Processing*, vol. 29, pp. 9032–9043, 2020.

[55] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[56] Y. Zhang and Q. Yang, "A survey on multi-task learning," *IEEE Transactions on Knowledge and Data Engineering*, 2021.

[57] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest *et al.*, "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, 2014.
[58] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. S. Kirby, J. B. Freymann, K. Farahani, and C. Davatzikos, "Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features," *Scientific Data*, vol. 4, no. 1, pp. 1–13, 2017.
[59] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, R. T. Shinohara, C. Berger, S. M. Ha, M. Rozycki *et al.*, "Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge," *arXiv preprint arXiv:1811.02629*, 2018.
[60] S. M. Smith, "Fast robust automated brain extraction," *Human brain mapping*, vol. 17, no. 3, pp. 143–155, 2002.
[61] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
[62] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in MRI images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1240–1251, 2016.
[63] Y. Ding, L. Gong, M. Zhang, C. Li, and Z. Qin, "A multi-path adaptive fusion network for multimodal brain tumor segmentation," *Neurocomputing*, vol. 412, pp. 19–30, 2020.
[64] M. Akil, R. Saouli, R. Kachouri *et al.*, "Fully automatic brain tumor segmentation with deep learning-based selective attention using overlapping patches and multi-class weighted cross-entropy," *Medical Image Analysis*, vol. 63, p. 101692, 2020.
[65] K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Medical Image Analysis*, vol. 36, pp. 61–78, 2017.
[66] F. Isensee, P. F. Jäger, P. M. Full, P. Vollmuth, and K. H. Maier-Hein, "nnU-net for brain tumor segmentation," in *International MICCAI Brainlesion Workshop*, 2020, pp. 118–132.
[67] T. Henry, A. Carre, M. Lerousseau, T. Estienne, C. Robert, N. Paragios, and E. Deutsch, "Brain tumor segmentation with self-ensembled, deeply-supervised 3D U-net neural networks: a BraTS 2020 challenge solution," *arXiv preprint arXiv:2011.01045*, 2020.
[68] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *International Conference on Computer Vision*, 2017, pp. 2961–2969.
[69] J. W. Johnson, "Adapting mask-rcnn for automatic nucleus segmentation," *arXiv preprint arXiv:1805.00500*, 2018.
[70] B. L. Welch, "The generalization of 'student's'problem when several different population varlances are involved," *Biometrika*, vol. 34, no. 1-2, pp. 28–35, 1947.

Prasun Chandra Tripathi recieved B.Tech. degree from Uttar Pradesh Technical University, Lucknow, and M.Tech. degree from University of Hyderabad, Hyderabad. He earned Ph.D. degree from Indian Institute of Technology (ISM), Dhanbad, India. Currently, he is working as a Postdoc researcher with Department of Computer Science, University of Sheffield, UK. His area of interests lie in medical image analysis, deep learning, and image processing. He is a student member of IEEE.

Soumen Bag received B.E. and M.Tech. degrees in computer science and engineering from NIT Durgapur, in 2003 and 2008, respectively, and the Ph.D. degree from IIT Kharagpur, in 2013. He is currently working as an Associate Professor with the Department of Computer Science and Engineering, IIT (ISM) Dhanbad. Prior to this, he worked at BCET Durgapur, India, and IIIT Bhubaneswar, India. He is the author of several reputed international journals and conferences. His research interests include the areas of medical image analysis, document image analysis, OCR for Indian scripts, handwritten forgery, and digital image forgery. He is a member of different international societies, like IEEE and IUPRAI. He was a recipient of the Institute Gold medal for holding the position of First-Class-First in his master's degree. He was also a recipient of different fellowships/scholarships from the national and international societies, like IEEE Signal Processing Society, CSIR, Government of India, Microsoft India, and MHRD, Government of India. He was a recipient of the Early Career Research Award (ECR) from SERB, Government of India, in 2017. He is enlisted in the Marquis Who's Who in the World, USA (32nd Edition, 2015). He acts as an organizing and programme committee member of different national and international conferences. He is currently acting as an Associate Editor of IET Image Processing journal.