Brain tumor segmentation and grading of lower-grade glioma using deep learning in MRI images

Mohamed A. Naser [a,*], M. Jamal Deen [b]

[a] *Department of Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, 77030, USA*
[b] *Department of Electrical and Computer Engineering and the School of Biomedical Engineering, McMaster University, Hamilton, Ontario, L8S 4K1, Canada*

A B S T R A C T

Gliomas are the most common malignant brain tumors with different grades that highly determine the rate of survival in patients. Tumor segmentation and grading using magnetic resonance imaging (MRI) are common and essential for diagnosis and treatment planning. To achieve this clinical need, a deep learning approach that combines convolutional neural networks (CNN) based on the U-net for tumor segmentation and transfer learning based on a pre-trained convolution-base of Vgg16 and a fully connected classifier for tumor grading was developed. The segmentation and grading models use the same pipeline of T1-precontrast, fluid attenuated inversion recovery (FLAIR), and T1-postcontrast MRI images of 110 patients of lower-grade glioma (LGG) for training and evaluations. The mean dice similarity coefficient (DSC) and tumor detection accuracy achieved by the segmentation model are 0.84 and 0.92, respectively. The grading model classifies LGG into grade II and grade III with accuracy, sensitivity, and specificity of 0.89, 0.87, and 0.92, respectively at the MRI images' level and 0.95, 0.97, and 0.98 at the patients' level. This work demonstrates the potential of using deep learning in MRI images to provide a non-invasive tool for simultaneous and automated tumor segmentation, detection, and grading of LGG for clinical applications.

1. Introduction

GLIOMAS are the most common brain tumors [1] with different grades according to tumor malignancy and the rate of growth [2]. Based on the World Health Organization (WHO), gliomas are classified into four grades: grade I which is usually curable by surgical resection; grade IV, Glioblastoma Multiform (GBM) which is the most aggressive with the lowest survival rate; and lower-grade glioma (LGG), grade II and grade III which are less aggressive and infiltrative than grade IV [2,3]. There are different histological types of LGG such as astrocytoma, oligoastrocytoma, and oligodendroglioma. Gliomas' grades and tumor location, shape, and size are highly required information to determine the survival and treatment plans which include surgery, chemotherapy, radiotherapy, or a combination [4]. Therefore, developing non-invasive techniques that automate tumor segmentation and grading would be beneficial for clinical use in diagnosis and adaptive treatment planning for optimizing patients' outcomes.

MRI is non-invasive and can provide detailed images of the brain. Therefore, it is commonly used for diagnosis and initial characterization of tumors. Brain tumor segmentation using MRI can be used to provide information about the tumor, such as its shape and size, in addition to tumor location for biopsy sampling that is used for tumor grading by a pathologist.

The gold standard for tumor segmentation is manual segmentation. However, it is costly, time-consuming, and suffers from inter-observer variability [5]. Therefore, automatic tumor segmentation is preferred, especially for large amounts of data and when continuous monitoring of tumor and adaptive treatment planning is required. However, accurate automatic tumor segmentation is usually challenging due to high variability of tumor locations, shapes and structures. Tumor grading using a pathology examination requires biopsies, so it is an invasive time-consuming technique and can suffer from inaccurate sampling and inter-observer variability [6,7]. Thus, automatic tumor segmentation and grading that are directly obtained from MRI would be highly beneficial for clinical use.

There are several approaches for automatic tumor segmentation using MRI that are mainly based on generative or discriminative models [5,8,9]. Generative models such as atlas-based models [10,11] require prior knowledge of anatomy and use posterior probabilities for voxels' classification accompanied by image registration for tumor

---

* Corresponding author.
  E-mail addresses: manaser@mdanderson.org (M.A. Naser), jamal@mcmaster.ca (M.J. Deen).

segmentation [11,12]. On the other hand, discriminative models such as machine learning-based Support Vector Machine (SVM) [13] and random forests [14] do not require prior knowledge of anatomy and use imaging features extracted from MRI rather than the original MRI data for voxels' classification. Dimensionality reduction and imaging features selection are usually implemented before model training. Deep learning using CNN is a promising machine learning approach that is different from discriminative models in which imaging feature extraction and selection is automated during model training [15,16] and has shown promise in automatic tumor segmentation [17–19].

Similar to tumor segmentation, machine learning approaches such as SVM [20] were used for tumor grading of gliomas using MRI. Deep learning models using CNN show promising results of grading gliomas using both digital pathology images [21] and MRI images [22]. However, for full automation of tumor grading, the selection of MRI images for tumors is required to be automated and independent of a physician's supervision. Therefore, for fully automated grading of gliomas, tumor segmentation and detection should precede tumor grading for MRI images' selection.

In this work, we developed combined deep learning segmentation and grading models based on CNN to provide simultaneous tumor segmentation, detection, and grading of LGG using MRI images. The predicted tumor masks by the segmentation model were used to automate the selection of tumor MRI images for grading. The results of this work demonstrate that MRI images of patients can be used in one pipeline by the segmentation and grading models to provide predictions for tumor segmentations as well as LGG tumor grades at both MRI images' level and patients' level. Therefore, the proposed approach can be used as a non-invasive and automated tool that could be helpful for diagnosis and treatment planning of LGG.

2. Methods

An overview of the proposed approach used for tumor segmentation and grading is summarized in Fig. 1. In brief, the brain MRI images and the corresponding tumor masks generated using manual segmentation were processed (2.1.3) and then used for training the segmentation model (2.2.1). The trained segmentation model and the processed MRI images were then used to automatically generate tumor masks (2.2.2). The grading model (2.3.1) used a sub-set of the processed MRI images based on the generated tumor masks for classifying images with different grades (2.3.2).
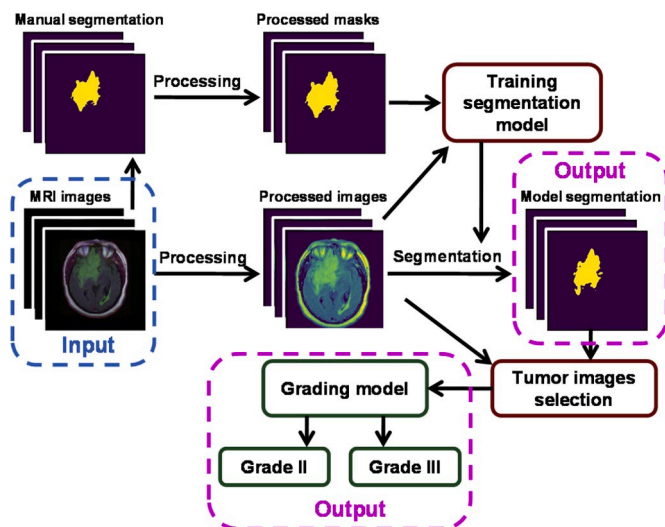

Fig. 1. An illustration of the workflow used for tumor segmentation and grading.

2.1. Dataset

2.1.1. Patients' data

The data used in this study is for a collection of 110 patients of LGG and is available at The Cancer Imaging Archive (TCIA) [23]. The group of patients is from 5 different institutions (Thomas Jefferson University – 16 patients, Henry Ford Hospital – 45 patients, UNC – 1 patient, Case Western – 14 patients, and Case Western St. Joseph's – 34 patients). The patients are classified as 50 patients with Grade II, 58 patients with Grade III, 2 patients with unknown's tumor grades, and 1 patient with unknown tumor sub-type. The patients' data that includes tumor grades, tumor sub-types, genders, and ages are given in Table 1.

2.1.2. Imaging data

The imaging data for the 110 patients are preoperative and are given for T1 pre-contrast, FLAIR, and T1 post-contrast MRI sequences. The manual segmentations and the corresponding generated tumor masks were performed by Buda et al. [17] using the FLAIR MRI images. The MRI images and the corresponding manually segmented tumor masks are given in TIF format images and made publically available for download from Ref. [24] by Buda et al. [17]. The 3 RGB channels of the TIF MRI images contain the pre-contrast, FLAIR and post-contrast MRI sequences, respectively. There are 6 patients with missing pre-contrast and 9 patients with missing post-contrast sequences. For those patients, the corresponding images' TIF channels use the FLAIR sequences instead to fill in the gaps. The MRI images per patient vary between 20 and 88 with size of 256×256 pixels and show cross-sectional areas of the brain as shown in Fig. 1. The total number of images, images with tumor, and images without tumor (normal tissue) from all patients are 3929, 1373, and 2556, respectively. Some MRI images at the beginning and/or the end of patients' scans have all pixels with zero values for some or all image channels. In addition, the images' level intensity varies between different patients. Therefore, a perquisite step for image processing (2.1.3) is required for the input data to the segmentation (2.2.1) and grading model (2.3.1).

2.1.3. Image processing

The first step is to crop unnecessary regions that correspond to non-tissue areas from the MRI and the corresponding tumor mask images for all images of each patient. The cropping was performed in three dimensions of images per patient (i.e., height, width, and elevation – image slices). After cropping, the cropped images are padded with zeros to retain the images' aspect ratio, and then the images are resized back to 256×256 pixels. The cropping, padding, and resizing were performed for both MRI and tumor mask images. Then, all MRI images intensity level was rescaled to be between 0 and 255 and finally the rescaled images were normalized such that all tissue regions pixel values were between −1 and 1, while the non-tissue regions were kept at −1 for all images as shown in Fig. 1. The non-zero pixel values of all tumor masks equal 1. The total number of images, tumor images, and normal tissue images from all patients after processing were 3121, 1373, and

Table 1
Lower-grade glioma tumor (LGG) and patient characteristics.

	Grade II N = 50	Grade III N = 58	Unknown N = 2
Tumor sub-types			
Astrocytoma	8	26	–
Oligoastrocytoma	14	14	1
Oligodendroglioma	28	18	–
Unknown	–	–	1
Gender			
Male	23	30	1
Female	27	28	–
Unknown	–	–	1
Age (years)	Mean	Std	Range
	46	14	20–75

1748, respectively.

2.2. Tumor segmentation

2.2.1. Segmentation model

The segmentation model used in this work was based on CNN with the U-net architecture [25]. As shown in Fig. 2A, the U-net consists of 4 blocks of layers which contain 2 convolution layers (red) with ReLU activation function and 1 max pooling layer (blue) in the encoding (down-sampling) part and a similar 4 blocks of layers but with 1 convolution transpose layer (orange) instead of max pooling in the decoding (up-sampling) part [26–29]. Convolution layers from the down-sampling and up-sampling parts are concatenated as shown in Fig. 1A. The number of filter channels is given at the top of each layer while the image size for both height and width is given at the bottom of each layer. The size of the input layer (green) is $256 \times 256 \times 1$ and the size of the output layer is $256 \times 256 \times 1$ which is a convolution layer with Sigmoid activation function. The U-net model was built using Keras with Tensor flow as backend [30,31].

2.2.2. Model implementation

The FLAIR channels of the processed MRI images were used as inputs to the segmentation model. The processed manual segmentation tumor masks were used as the ground truth for the model. The processed images and masks were split into training and test data (2.2.4) and then used for training and validating the segmentation model. The optimizer used was Adam with a learning rate equals $5*10^{-5}$. The model metrics was the dice similarity coefficient (DSC), Eq. (2), and the loss function was the negative value of DSC. The batch size and the number of epochs were 20 and 100, respectively. To overcome over-fitting, data augmentation of the processed MRI and masks' images was implemented using a rotation range of 25°, and zoom, shift and shear ranges of 20%, and enabled horizontal-flip.

The number of normal tissue images is larger than that of tumor images which makes the two classes of images unbalanced. To reduce the unbalancing and improve the prediction of tumor masks in images with tumor, we scaled the loss function by a weight that is a function of the number of pixels in the provided ground truth manual segmentation mask, as shown in Eq. (1),

$$\mathbf{w}_i^n = N^n * \frac{\sum_{j,k} \mathbf{M}_{i,j,k}^n + 0.1*\sum_{i,j,k} \mathbf{M}_{i,j,k}^n \big/ N^n}{1.1*\sum_{i,j,k} \mathbf{M}_{i,j,k}^n}, \tag{1}$$

where $\mathbf{w}_i^n$ is the sample-weight for image i of patient n, and $\mathbf{M}_{i,j,k}^n$ is the value of ground-truth mask image i of patient n at the pixel (j,k) and N^n is the number of MRI images of patient n. The summation over i is from 1 to N^n and over j and k is from 1 to 256. Using Eq. (1), the weight of the loss function that corresponds to tumor with larger cross-sectional area will be larger than that with smaller areas as well as normal tissue images.

2.2.3. Tumor detection

The segmentation model can identify tumor in the MRI images at the pixel level by providing predictions for tumor masks that show the location, shape and size of the tumor. At the image level, a tumor detection approach has been developed to classify the MRI images into normal tissue or tumor using the corresponding predicted tumor masks by the segmentation model. In this approach, the MRI image is classified normal tissue if the corresponding tumor mask has zero values for all image pixels and tumor if the corresponding tumor mask has image pixels of non-zero values. Therefore, the segmentation model outcomes were used for tumor detection at both the pixel and image levels.

2.2.4. Model validation

There is no available separate data that can be used to evaluate and validate the performance of the segmentation model. Therefore, we used a cross-validation approach where the 110 patients' imaging data (2.1.1) and (2.1.2) and the corresponding ground truth tumor masks
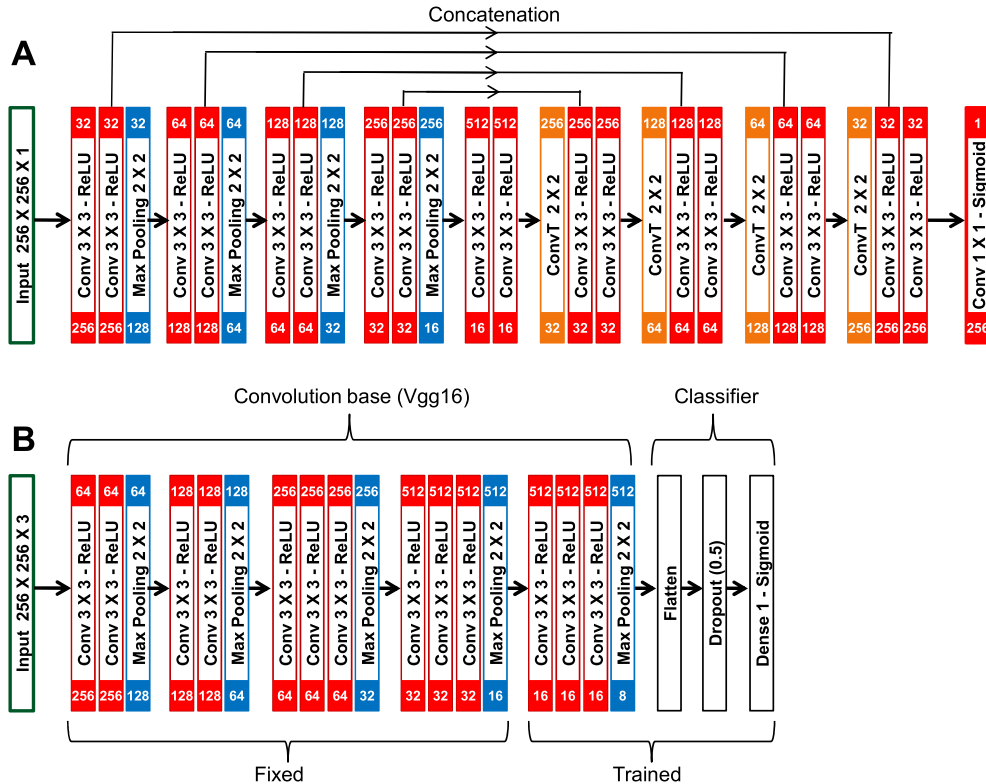

Fig. 2. Illustrations of (A) the segmentation model (U-net architecture) and (B) the grading model (a densely connected neural network classifier on top of the Vgg16 convolution-base).

were split into 5 sets (Set 1 to Set 5). Each set contains imaging data of 22 patients randomly selected from the 110 patients' dataset. Each set of 22 patients serves one time as test data for the segmentation model trained using imaging data from the remaining 4 sets. Using this approach, the segmentation model was trained and tested 5 times and then the average DSC value was obtained using the average DSC values obtained for the test data sets using the corresponding trained segmentation models.

2.3. Tumor grading

2.3.1. Grading model

The grading model was based on transfer learning where a classifier was built on top of a pre-trained convolution base of the Vgg16 network [30] as shown in Fig. 2B. The weights of the pre-trained convolution base of Vgg16 can be downloaded from Ref. [32]. The convolution base consists of 5 blocks of convolution layers with ReLU activation function and a max pooling layer. A classifier that consists of a flatten layer for converting data from 3D to 1D, a Dropout layer to overcome over fitting, and a densely connected layer with the Sigmoid activation function as the output layer. The size of the input and output layers are $256 \times 256 \times 3$ and 1, respectively. The grading model was built-up using Keras with Tensor flow as backend [30].

2.3.2. Model implementation

The segmentation model was used to provide predicted tumor masks using the FLAIR channels of the MRI images. These masks were used to exclude normal tissue images (2.2.3) and to select representative MRI images for tumor grading. The tumor MRI images for each patient were selected such that the tumor cross-sectional areas, estimated from the corresponding predicted tumor masks, were with 50% of the largest tumor cross-sectional area in the whole patient MRI scan. This constraint ensures excluding false positive predictions of tumors from the segmentation model and only includes images with relevant grading information about tumor such as necrotic cores, tumor shape irregularity, and tumor size. These information are usually pronounced close to the tumor center and less pronounced close to the tumor edges.

The selected tumor MRI images (815 images – 400 grade II and 415 grade III) with image size equals $256 \times 256 \times 3$ were split into training and test data (2.3.3) and then used for training and validating the grading model. The optimizer used was Adam with a learning rate equals $5*10^{-5}$, the model metrics was binary accuracy, and the loss function was binary cross-entropy. The batch size and the number of epochs were 10 and 50, respectively. Data augmentation similar to that used in (2.2.2) was used but with zoom, shift and shear ranges of 40% to overcome over-fitting using a smaller number of imaging data.

After training, the grading model was used for predicting the tumor grades of the selected MRI images for each patient. Then, the predicted grades of the selected MRI images were used to determine the tumor grade for each patient based on a majority vote of the patient selected MRI tumor images.

2.3.3. Model validation

The 400 and 415 images of tumor grades II and III, respectively were randomized and split into 5 sets (S1 to S5). Each set contains approximately similar number of grade II and grade III tumors like that in other sets. Using cross-validation, each set was used to test the model trained using imaging data from the remaining 4 sets. The grading model was trained and tested 5 times and then the average binary-accuracy was obtained using the average binary-accuracy values obtained for the test data sets using the corresponding trained grading model.

2.4. Evaluation

The DSC was used to evaluate the similarity of the predicted tumor masks by the segmentation model (2.2.1) with the tumor masks obtained by manual segmentation. For manual and model segmentation

masks $\mathbf{M}^{manual}$ and $\mathbf{M}^{model}$, respectively, the DSC is defined as

$$DSC = \frac{\sum_{j,k} \mathbf{M}^{manual}_{j,k} \mathbf{M}^{model}_{j,k} + 1}{\sum_{j,k} \mathbf{M}^{manual}_{j,k} + \sum_{j,k} \mathbf{M}^{model}_{j,k} + 1}, \tag{2}$$

where the summation over j and k is for all image pixels.

To evaluate the tumor detection (2.2.3) by the segmentation model, we used the following metrics: accuracy, sensitivity, and specificity, which are defined as:

$$accuracy = \frac{\text{True predicted normal and tumor images}}{\text{All predicted images}}. \tag{3}$$

$$sensitivity = \frac{\text{True predicted tumor images}}{\text{All observed tumor images}}. \tag{4}$$

$$specificity = \frac{\text{True predicted normal images}}{\text{All observed normal images}}. \tag{5}$$

The metrics used for evaluating the grading model were accuracy, sensitivity, and specificity, which are defined as:

$$accuracy = \frac{\text{True predicted tumor grades}}{\text{All observed tumor grades}}. \tag{6}$$

$$sensitivity = \frac{\text{True predicted tumor grades III}}{\text{All observed tumor grades III}}. \tag{7}$$

$$specificity = \frac{\text{True predicted tumor grades II}}{\text{All observed tumor grades II}}. \tag{8}$$

3. Results

3.1. Tumor segmentation

The DSC values' distributions obtained by the segmentation model for the 5 test data sets are illustrated in Fig. 3A. The DSC median values for Set 1 to Set 5 are 0.95, 1, 0.97, 1 and 1, respectively, while the DSC mean values are 0.80, 0.84, 0.82, 0.87 and 0.86, respectively. The DSC median and mean values for all test data sets are 0.98 and 0.84, respectively, with standard deviations of 0.02 and 0.03, respectively.

The confusion matrix for tumor detection results aggregated using all the test data sets and obtained by the segmentation model is shown in Fig. 3B. The classification accuracy, tumor detection sensitivity and normal tissue specificity are 0.92 for all of them, as summarized in Table 2.

To illustrate the performance of the segmentation model, samples of heatmaps and overlays of FLAIR MRI images with the outlines of tumor masks using manual and model segmentations for the test data sets are shown in Fig. 4. At the left of each panel in Fig. 4, a heat map image shows the highlighted tumor in pink, while the overlay image with tumor outlines (red – manual segmentation and green – model segmentation) is given at the right. The tumor grade and the DSC value are given at the top and the top right corner of each panel. The first, second, and third rows show representative segmentation results for DSC values below, similar, and above the segmentation model's mean DSC value of 0.84.

The confusion matrix for tumor grading results aggregated using all MRI images of the test data sets S1 to S5 and obtained by the grading model is shown in Fig. 5A. The classification accuracy, grade III tumor detection sensitivity, and grade II tumor detection specificity are 0.89, 0.87, and 0.92, respectively, as shown in Table 2. The Receiver Operating Characteristics (ROC) curves for the grading model using the 5 test data sets S1 to S5 are given in Fig. 5B. The areas under the curves (AUCs) for S1 to S5 are 0.98, 0.96, 0.97, 0.96, and 0.97, respectively, and the mean AUC is 0.97.
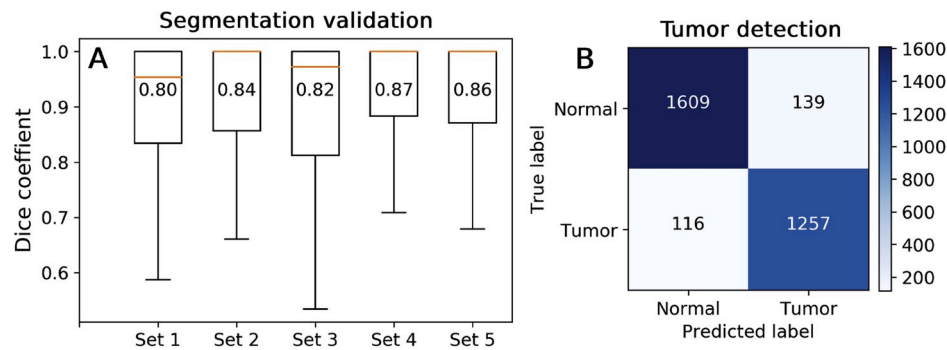
Fig. 3. (A) Boxplots of the DSC distribution for the 5 test data sets (Set 1 to Set 5) used for segmentation model validation. The DSC mean values are given in the boxes and the orange lines refer to the DSC median values. (B) Confusion matrix for the tumor detection showing the number of images of the true and predicted normal and tumor labels. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 2
Tumor detection and grading results.

	Tumor detection	Tumor grading		
		Images	Patient (model)	Patient (manual)
Accuracy	0.92	0.89	0.95	0.95
Sensitivity	0.92	0.87	0.93	0.97
Specificity	0.92	0.92	0.98	0.98

3.2. Tumor grading

At the patient level, the confusion matrices for tumor grading results aggregated using 108 patients' data with known grading using MRI images selection based on model and manual segmentations are given in Fig. 5C and D, respectively. The tumor grading accuracy and grade II tumor specificity are 0.95 and 0.98, respectively for both model and manual segmentations. The grade III tumor sensitivity for model and manual segmentations are 0.93 and 0.97, respectively. The grading results are summarized in Table 2.

Fig. 6 shows samples of the tumor grading results for two patients with tumors of grade II (Fig. 6A) and grade III (Fig. 6B). The grading model classifies 4 out of 6 images with grade II (Fig. 6A) and 5 out of 6 images with grade III (Fig. 6B). Based on majority votes, patient in Fig. 6A is classified with grade II tumor and patient in Fig. 6B is classified with grade III tumor.
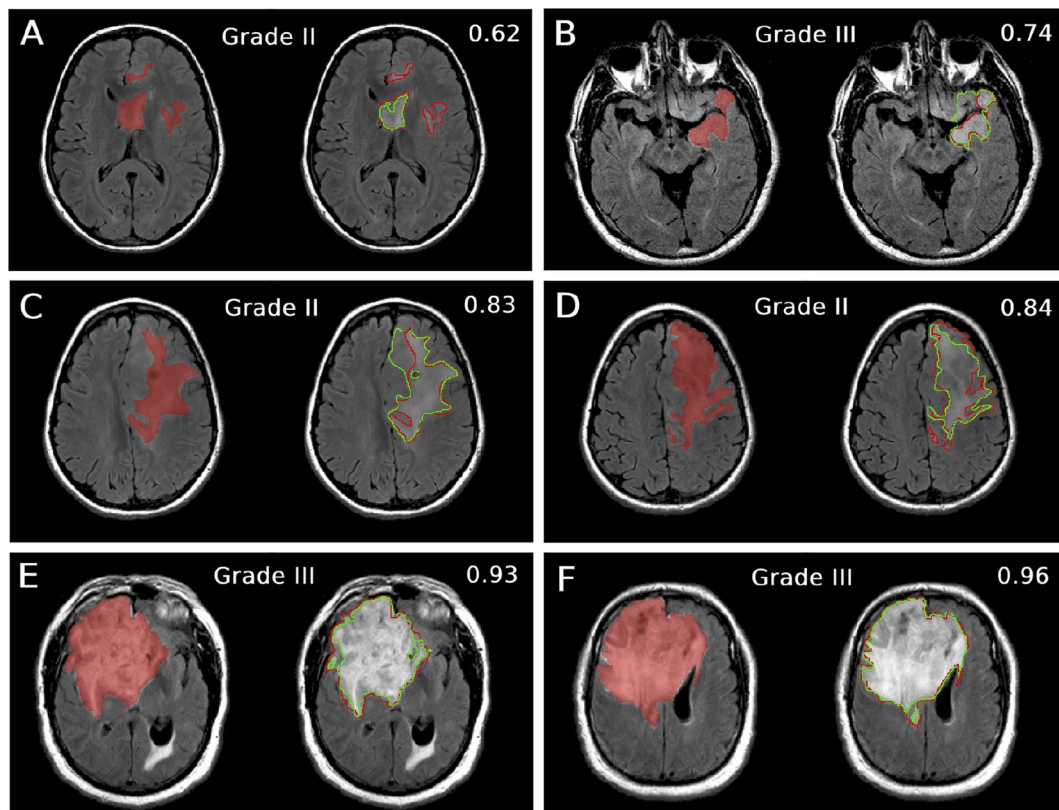

Fig. 4. Samples from the test data sets showing heatmaps of FLAIR images for highlighted tumor regions and overlays of FLAIR images and tumor masks' outlines (red – manual segmentation and green – model segmentation) for DSC values (A) and (B) lower than, (C) and (D) similar, and (E) and (F) larger than model DSC mean value (0.84). The heat maps are shown at the left while the overlays are shown at the right of each panel. The tumor grade and the DSC value are given at the top and top right of each panel. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)
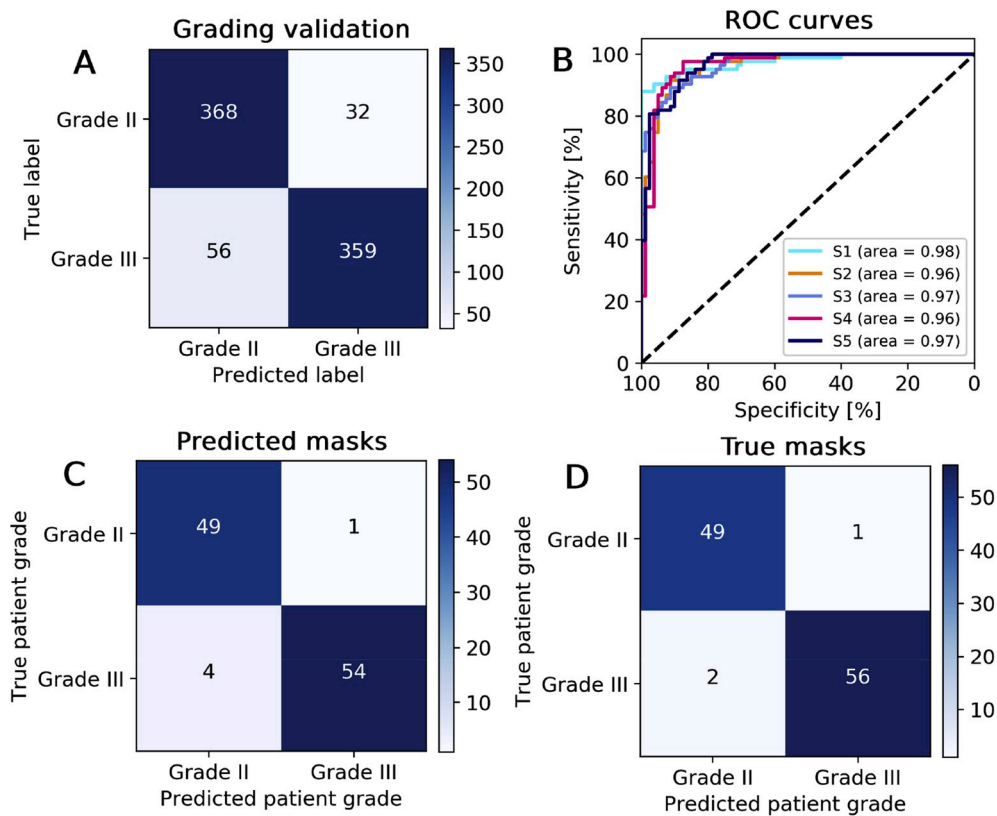
Fig. 5. (A) Confusion matrix for the grading models showing the number of images of the true and predicted tumor grades, (B) Receiver Operating Characteristics (ROC) curves for the grading model using the 5 different test data sets S1 – S5. The area under the curves are given in the figure label, (C), and (D) Confusions matrix for the patient tumor grading using MRI images selection based on predicted tumor masks (model segmentation) and true tumor masks (manual segmentation), respectively.
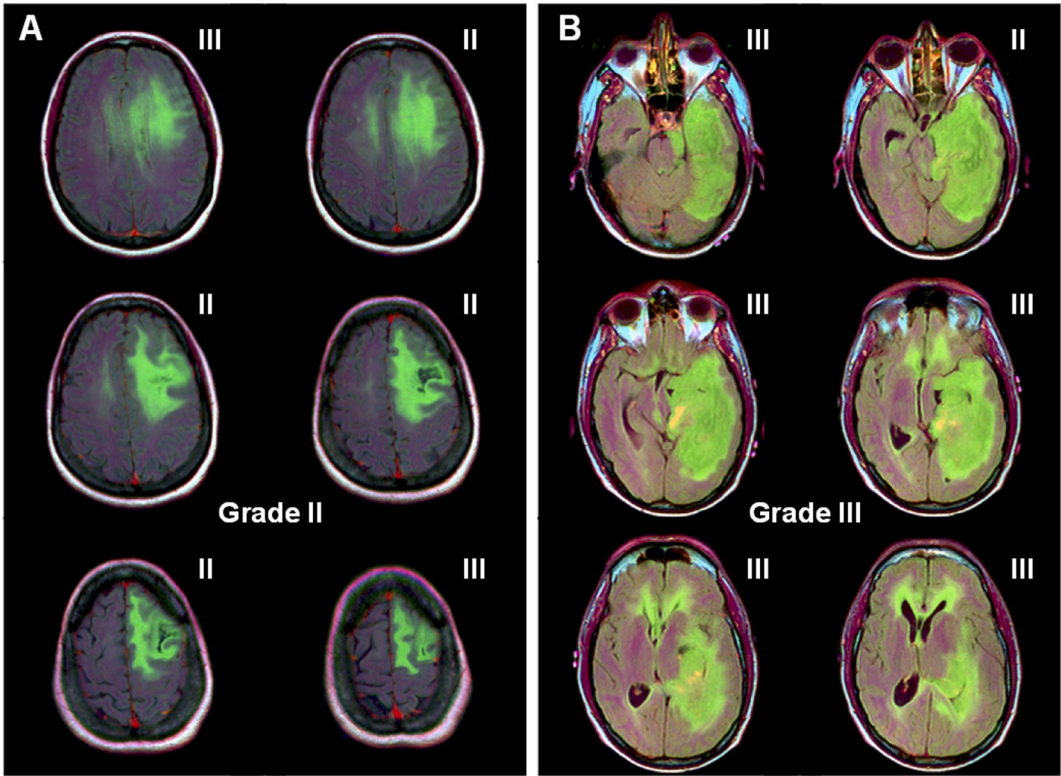

Fig. 6. (A) A sample patient classified with grade II tumor (4 out of six images classified with grade II tumor). (B) A sample patient classified with grade III tumor (5 out of six images classified with grade III tumor).

4. Discussion

In this work, we show that deep learning models based on CNN using the U-net and transfer learning using a pre-trained convolution-base of Vgg16 can be used for simultaneous tumor segmentation, detection and grading of LGG using MRI. The same pipeline of LGG MRI is used by the segmentation and grading models and the mean tumor segmentation DSC value, tumor detection accuracy, and tumor grading accuracy at image and patient levels obtained are 0.84, 0.92, 0.89 and 0.95, respectively.

The mean DSC value of 0.84 with standard deviation of 0.03 obtained by the current segmentation model (Fig. 3A) is similar to the DSC value of tumor masks obtained using manual segmentations produced by two expert raters (0.84 with 0.02 standard deviations) [5]. The segmentation model used only the FLAIR MRI as a single channel input since the manual segmentation tumor masks were produced using the FLAIR MRI. In addition, the pre- and post-contrast MRI channels do not show pronounced tumor enhancement. Therefore, there are no relevant information to be added by the pre- and post-contrast MRI for tumor segmentation. Using different sample-weights during training the segmentation model, Eq. (1), leads to reducing the effect of the un-balance between normal and tumor MRI and therefore improving the overall mean DSC value, especially for MRI images with large tumor cross-sectional areas. In a previous study [17], a deep learning U-net model was applied to the same imaging data used in the current study. The reported mean DSC value using 1 out 22 cross-validation and the 3 RGB channels of MRI (pre-contrast, FLAIR, and post-contrast) was 0.82. The CNN-based model in Ref. [33] reported DSC value of 0.88 using BRATS 2013 and 0.78 using BRATS 2015 MRI data for complete tumor segmentation with an average of 0.83 considering both data sets. So, the reported 0.84 DSC value by the current model is higher or similar to some comparable deep learning models reported for brain tumor segmentation.

As shown in Fig. 3B, the predicted tumor masks can be used to classify MRI images to normal or tumor. The classification accuracy, sensitivity, and specificity are 0.92 for each. In Ref. [34], using a random forest classification method applied to MR spectroscopy data, rather than FLAIR MRI as in the current study, the reported classification sensitivity of distinguishing tumor from benign was 0.86 with AUC of 0.91 for the classifier. The classification accuracy and sensitivity of distinguishing normal tissue, LGG and higher-grade glioma using FLAIR MRI and CNN-based classification model reported in Ref. [22] were 0.91 and 0.92, respectively, which are comparable to the tumor detection results obtained by the proposed segmentation model.

The grading model used the 3 RGB channels of MR images, rather than just the FLAIR MRI channel used by the segmentation model. Due to the relatively small number of images, 815, used by the grading model compared to 3121 used by the segmentation model, the cross-validation results for the classifier as shown in Fig. 5A and B are given at the MRI images' level rather than patients' level as implemented in the segmentation model. The cross-validation approach for the grading model based on 1 out 5 data sets (i.e., 80% of data for training and 20% of data for testing – repeated 5 times) was considered to increase the statistical representation of the training data to the whole data for each of the 5 models used for the cross-validation. In another approach, different combinations of the data sets can be used to increase the number of train-test data sets available and therefore, improves the estimation of the model accuracy. For example, considering all different combinations of 2 out of 5 data sets (i.e., 60% of data for training and 40% of data for testing) will result in 10 different combinations of train-test data sets. However, the benefit of increasing the number of train-test data sets for the estimation of the model accuracy could be mitigated by the reduction of the statistical representation of the training data used by each model, especially for a relatively small data size. Therefore, the 1 out of 5 cross-validation used is considered a tradeoff between the size of the training data and the number of train-test data sets used for estimating

the model accuracy. The 815 tumor images were selected such that the tumor cross-sectional areas are within 50% of the largest tumor cross-sectional area in the whole scan. Increasing the threshold above 50% will result in providing images with more relevant imaging features for grading (i.e., excluding false positives and images with tiny tumors). However, on the other hand, it will reduce the available images for training. So, the choice of 50% was found to be a good compromise. The reported grading accuracy of LGG (distinguishing gliomas grade II from grade III) using digital pathology images and a CNN-based classifier was 0.71 [21]. The reported accuracy and sensitivity of classifying gliomas of high grade and LGG using FLAIR MRI and CNN-based model [22]; and T1 and FLAIR MRI and SVM-based classifier [20] were 0.91, 0.92; and 0.88 and 0.85, respectively. The LGG tumor grading accuracy, sensitivity and specificity obtained by the proposed grading model as shown in Table 2 and Fig. 5 are 0.89, 0.87, and 0.92, respectively at the image level, and 0.95, 0.93, and 0.98, respectively at the patient level. The reasons for the improved grading performance at the patient level is that considering the majority vote for grading the tumor at the patient levels will overcome some false grading at the image level as illustrated in Fig. 6. The performance of the proposed grading model at the image level is comparable with that reported in Ref. [20,22], though the grading was between high grade and LGG and not between gliomas grade II and grade III, and significantly higher than 0.71 grading accuracy reported in Ref. [21] using digital pathology images. A summary of comparisons between the proposed models' results and that of previous models mentioned above is given in Table 3.

There are some concerns regarding the proposed models that could be addressed for future improvements. The LGG data used for validation is relatively small and there was no independent data set available for testing. Therefore, a cross-validation approach was adopted to evaluate the models' performance. However, for generalization of the models, especially for the grading model where the overall number of images is relatively small, additional data sets should be used for more accurate evaluation and to allow cross-validation at the patient level using data from multiple institutions. For the segmentation model, the mean DSC value obtained by the proposed U-net model was comparable with the DSC value expected using manual segmentations due to inter-observer variability. However, there are several other network architectures that were proposed for tumor segmentations such as ResNet [35], Inception [36], and DenseNet [37]. Therefore, it would be interesting to investigate and compare results from these network architectures with the proposed approach. In addition, the proposed transfer learning used a relatively simple network architecture based on Vgg16, which is suitable for the size of the data set in the current study for training and validation. However, for a larger data set that contains wide range of gliomas MRI, other more advanced networks such as Vgg19, ResNet50,

Table 3

Summary of results' comparisons: DSC, accuracy (acc), Sensitivity (sen), and Specificity (spe).

	Model	Results	Data
Tumor Segmentation	Proposed U-net	DSC = 0.84	FLAIR MRI
	U-net [17]	DSC = 0.82	Pre-, FLAIR, Post-contrast MRI
	CNN [33]	DSC = 0.88 & 0.78	MRI BRATS 2013 & 2015
Tumor Detection	Proposed Tumor detection	ACC = SEN = SPE = 0.92	FLAIR MRI
	Random Forest [34]	SEN = 0.86	MR Spectroscopy
Tumor Grading	Proposed Vgg16	ACC = 0.89, SEN = 0.87, and SPE = 0.92	Pre-, FLAIR, Post-contrast MRI
	CNN [22]	ACC = 0.9 & SEN = 0.92	FLAIR MRI
	CNN [21]	ACC = 0.71	Digital pathology

Inception V3, and Xception [30] can be of interest for improving classification accuracy. Also, the grading model was used for classifying LGG into grade II and grade III. However, expanding the classification to include histological sub-types such as that shown in Table 1 and/or tumor genomic subtypes [17] might be of high interest for future clinical use.

5. Conclusions

This work illustrates the potential of using deep learning and transfer learning models in MRI images to provide accurate automatic tumor segmentation, detection and grading of LGG brain tumors. The use of one pipeline of MRI images by the segmentation and grading models allows full automation and simultaneous segmentation and grading of the brain tumors. Additionally, given a segmentation DSC mean value of 0.84, tumor detection accuracy of 0.92, and tumor grading accuracy at image and patient levels of 0.89 and 0.95, respectively, the proposed approach shows a promise as a non-invasive tool for tumor characterization in LGG.

Declaration of competing interest

'None Declared'.

Acknowledgments

References

[1] Q.T. Ostrom, et al., The Epidemiology of Glioma in Adults: A State of the Science Review, Neuro Oncol. 16 (7) (2014) 896–913.

[2] D.N. Louis, et al., The 2007 WHO classification of tumours of the central nervous system, Acta Neuropathol. 114 (2) (2007) 97–109.

[3] D.N. Louis, et al., The 2016 World Health organization classification of tumors of the central nervous system: a summary, Acta Neuropathol. 131 (6) (2016) 803–820.

[4] G. Tabatabai, et al., Molecular diagnostics of gliomas: the clinical perspective, Acta Neuropathol. 120 (5) (2010) 585–592.

[5] B.H. Menze, et al., The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS), IEEE Trans. Med. Imaging 34 (2015) 1993–2024, 10th ed.

[6] K.D. Miller, et al., "Cancer Treatment and Survivorship Statistics, 2016,", Cancer J. Clin. 66 (2016) 271–289, 4th ed.

[7] S.G. Bowden, et al., Sodium Fluorescein Facilitates Guided Sampling of Diagnostic Tumor Tissue in Nonenhancing Gliomas, Neurosurgery 82 (2018) 719–727, 5th ed.

[8] S. Bauer, R. Wiest, L.P. Nolte, M. Reyes, A survey of MRI-based medical image analysis for brain tumor studies, Phys. Med. Biol. 58 (13) (2013) R97–R129.

[9] E. Angelini, O. Clatz, E. Mandonnet, E. Konukoglu, L. Capelle, H. Duffau, Glioma Dynamics and Computational Models: A Review of Segmentation, Registration, and in Silico Growth Algorithms and Their Clinical Applications, Curr. Med. Imaging Rev. 3 (2007) 262–276, 4th ed.

[10] M. Prastawa, E. Bullitt, S. Ho, G. Gerig, A brain tumor segmentation framework based on outlier detection, Med. Image Anal. 8 (3) (2004) 275–283.

[11] M. Prastawa, E. Bullitt, S. Ho, G. Gerig, Robust estimation for brain tumor segmentation, in: In Lecture Notes in Computer Science, PART 2 2879, Springer, Berlin, Heidelberg, 2003, pp. 530–537.

[12] S. Parisot, H. Duffau, S. Chemouny, N. Paragios, Joint tumor segmentation and dense deformable registration of brain MR images, in: In Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 7511, Springer, Berlin, Heidelberg, 2012, pp. 651–658.

[13] S. Bauer, L.P. Nolte, M. Reyes, Fully automatic segmentation of brain tumor images using Support vector machine classification in combination with hierarchical conditional random field regularization, in: In Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 6893, Springer, Berlin, Heidelberg, 2011, pp. 354–361.

[14] D. Zikic, et al., Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel MR, in: In Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 7512, Springer, Berlin, Heidelberg, 2012, pp. 369–376.

[15] M.P. McBee, et al., "Deep Learning in Radiology,", Academic Radiology 25 (2018) 1472–1480, 11th ed.

[16] M.A. Mazurowski, M. Buda, A. Saha, M.R. Bashir, Deep learning in radiology: an overview of the concepts and a survey of the state of the art with focus on MRI, J. Magn. Reson. Imag. 49 (4) (2019) 939–954.

[17] M. Buda, A. Saha, M.A. Mazurowski, Association of genomic subtypes of lower-grade gliomas with shape features automatically extracted by a deep learning algorithm, Comput. Biol. Med. 109 (Jun. 2019) 218–225.

[18] M. Havaei, et al., Brain tumor segmentation with deep neural networks, Med. Image Anal. 35 (Jan. 2017) 18–31.

[19] M. Havaei, N. Guizard, H. Larochelle, P.M. Jodoin, Deep learning trends for focal brain pathology segmentation in MRI, in: In Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 9605, Springer, Cham, 2016, pp. 125–148.

[20] E.I. Zacharaki, et al., Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme, Magn. Reson. Med. 62 (6) (Dec. 2009) 1609–1618.

[21] M.G. Ertosun, D.L. Rubin, Automated Grading of Gliomas using Deep Learning in Digital Pathology Images: A modular approach with ensemble of convolutional neural networks, in: AMIA Annu. Symp. proceedings. AMIA Symp., vol. 2015, 2015, pp. 1899–1908.

[22] S. Khawaldeh, U. Pervaiz, A. Rafiq, R.S. Alkhawaldeh, Noninvasive grading of glioma tumor using magnetic resonance imaging with convolutional neural networks, Appl. Sci. 8 (1) (2017) 27.

[23] TCGA-LGG - the cancer imaging archive (TCIA) public access - cancer imaging archive wiki [Online]. Available, https://wiki.cancerimagingarchive.net/display/Public/TCGA-LGG.

[24] brain-segmentation-pytorch | Kaggle. [Online]. Available, https://www.kaggle.com/mateuszbuda/brain-segmentation-pytorch/data.

[25] O. Ronneberger, P. Fischer, T. Brox, "U-net: convolutional networks for biomedical image segmentation," in, Int. Conf. Med. Image Comput. Comput. Interv 9351 (2015) 234–241.

[26] H. Noh, S. Hong, B. Han, Learning deconvolution network for semantic segmentation, in: Proceedings of the IEEE International Conference on Computer Vision, ICCV, 2015, pp. 1520–1528.

[27] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR, 2015, pp. 3431–3440.

[28] V. Badrinarayanan, A. Kendall, R. Cipolla, SegNet: a deep convolutional encoder-decoder architecture for image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 39 (12) (2017) 2481–2495.

[29] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks, Commun. ACM 60 (2017), 6th ed.

[30] F. Chollet, Deep Learning with Python & Keras, vol. 80, Manning Publications Co., 2018, 1.

[31] X. Wang, L.T. Yang, J. Feng, X. Chen, M.J. Deen, A tensor-based big service framework for enhanced living environments, IEEE Cloud Comput. 3 (6) (Nov. 2016) 36–43.

[32] F. Chollet, "Vgg16_weights." [Online]. Available: https://github.com/fchollet/deep-learning-models/releases/download/v0.1/vgg16_weights_tf_dim_ordering_tf_kernels_notop.h5.

[33] S. Pereira, A. Pinto, V. Alves, C.A. Silva, Brain tumor segmentation using convolutional neural networks in MRI images, IEEE Trans. Med. Imag. 35 (5) (2016) 1240–1251.

[34] G. Ranjith, R. Parvathy, V. Vikas, K. Chandrasekharan, S. Nair, Machine Learning Methods for the Classification of Gliomas: Initial Results Using Features Extracted from MR Spectroscopy, Neuroradiol. J. 28 (2015) 106, 2nd ed.

[35] Q. Zhang, Z. Cui, X. Niu, S. Geng, Y. Qiao, Image segmentation with pyramid dilated convolution based on ResNet and U-net, in: In Lecture Notes in Computer Science 10635, Springer, Cham, 2017, pp. 364–372.

[36] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR, 2016, pp. 2818–2826.

[37] S. Jegou, M. Drozdzal, D. Vazquez, A. Romero, Y. Bengio, The one hundred layers tiramisu: fully convolutional DenseNets for semantic segmentation, in: In IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPRW, 2017, pp. 1175–1183.