



CNN-Based Fully Automatic Glioma Classification with Multi-modal Medical Images

Bingchao Zhao^{1,2}, Jia Huang¹, Changhong Liang¹, Zaiyi Liu¹,
and Chu Han¹(✉)

¹ Department of Radiology, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, Guangdong 510080, China

² The School of Computer Science and Engineering, South China University of Technology, Guangzhou, Guangdong 510006, China

Abstract. The accurate classification of gliomas is essential in clinical practice. It is valuable for clinical practitioners and patients to choose the appropriate management accordingly, promoting the development of personalized medicine. In the MICCAI 2020 Combined Radiology and Pathology Classification Challenge, 4 MRI sequences and a WSI image are provided for each patient. Participants are required to use the multi-modal images to predict the subtypes of glioma. In this paper, we proposed a fully automated pipeline for glioma classification. Our proposed model consists of two parts: feature extraction and feature fusion, which are respectively responsible for extracting representative features of images and making prediction. In specific, we proposed a segmentation-free self-supervised feature extraction network for 3D MRI volume. And a feature extraction model is designed for the H&E stained WSI by associating traditional image processing methods with convolutional neural network. Finally, we fuse the extracted features from multi-modal images and use a densely connected neural network to predict the final classification results. We evaluate the proposed model with F1-Score, Cohen's Kappa, and Balanced Accuracy on the validation set, which achieves 0.943, 0.903, and 0.889 respectively.

Keywords: Glioma classification · Convolutional neural networks · Multiple modalities

1 Introduction

As the most common primary malignant tumor of the central nervous system, glioma comprise approximately 100, 000 newly diagnosed cases each year [2]. According to the 2016 World Health Organization (WHO) classification, diffuse glioma is categorized into five subtypes based on both its tumor histology and molecular alterations, among which glioblastoma, astrocytoma and oligodendroglioma are further designated with genetic subgroups, including promoter mutations in TERT, IDH mutations and chromosome arms 1p and 19q co-deletion [14]. Glioblastoma accounts for 70–75% of all diagnoses in adults, and as

well has the poorest prognosis with a 5-year survival rate smaller than 5%. On the other hand, anaplastic astrocytoma and oligodendroglioma are rarer than glioblastoma, associated with relatively better overall survivals [1,9]. The treatment strategies, including adjuvant therapy selections after surgery, chemotherapy regimens and dosing schedules, vary by different subtypes [8]. Therefore, an accurate classification of glioma based on radiological and pathological images would be valuable for clinical practitioners and patients to choose the appropriate management accordingly, promoting the development of personalized medicine.

As a non-invasive clinical procedure, magnetic resonance image (MRI) reveals the characteristics of tumor phenotypes. Thus, it has been widely used for computer-aided diagnosis of glioma. Existing approaches usually require either manual or semi-automatic tumor segmentation before quantitatively analyzing 3D MRI volumes, which is labor intensive and time consuming. Besides considering a single MRI sequence, it is worth to associate multiple sequences due to the differences in signal intensities and patterns between different tissues on them. What's more, the enhancement patterns depending on various tumor subtypes would also greatly enrich the information of the tumors and their surrounding areas. Therefore, not only conventional non-enhanced T1-weighted and T2-weighted images, but also post-contrast images should be investigated.

Histopathology slide is the gold standard for the cancer diagnosis. It reflects the tumor microenvironment. The born of the digital whole slide scanner makes it possible for computers to quantitatively analyze diffuse gliomas at the microscopic level. However, even with the rapid development of computer hardware, the extreme resolution of whole slide image (WSI) is still the obstacle towards fully automatic clinical adoption.

In the real world scenario, it is meaningful and vital to read both radiological and pathological images for their diagnostic values in different subtypes of gliomas. However, manual quantification of multi-omics images is commonly subjective and experience dependent. Therefore, many researchers dedicate to find out an automatic and objective way to quantify multi-omics images. But how to fuse the features of the images acquired from different modals in a reasonable and interpretable manner is still unknown.

The CPM-RadPath 2020 MICCAI challenge is conducted for automatic brain tumor classification using two different modal images, including radiology and pathological images. Each case provides MRI of the four modalities of native (T1), post-contrast T1-weighted (T1Gd), T2-weighted (T2), and T2 Fluid Attenuated Inversion Recovery (T2-FLAIR), and an H&E stain digitized whole-slide images(WSI) which was scanned at 20x or 40x magnification. This challenge provides 388 glioma cases, including three types of gliomas, i.e., glioblastoma, oligodendroglioma, and astrocytoma, divided into a training set (70% of cases), validation set (20%), and test set (10%). In this paper, we proposed an effective pipeline for multi-modal images tumor classification. Firstly, a segmentation-free self-supervised feature extraction network is proposed for 3D MRI volume. Secondly, we proposed a feature extraction model for the H&E stained WSI by associating traditional image processing methods with convolutional neural

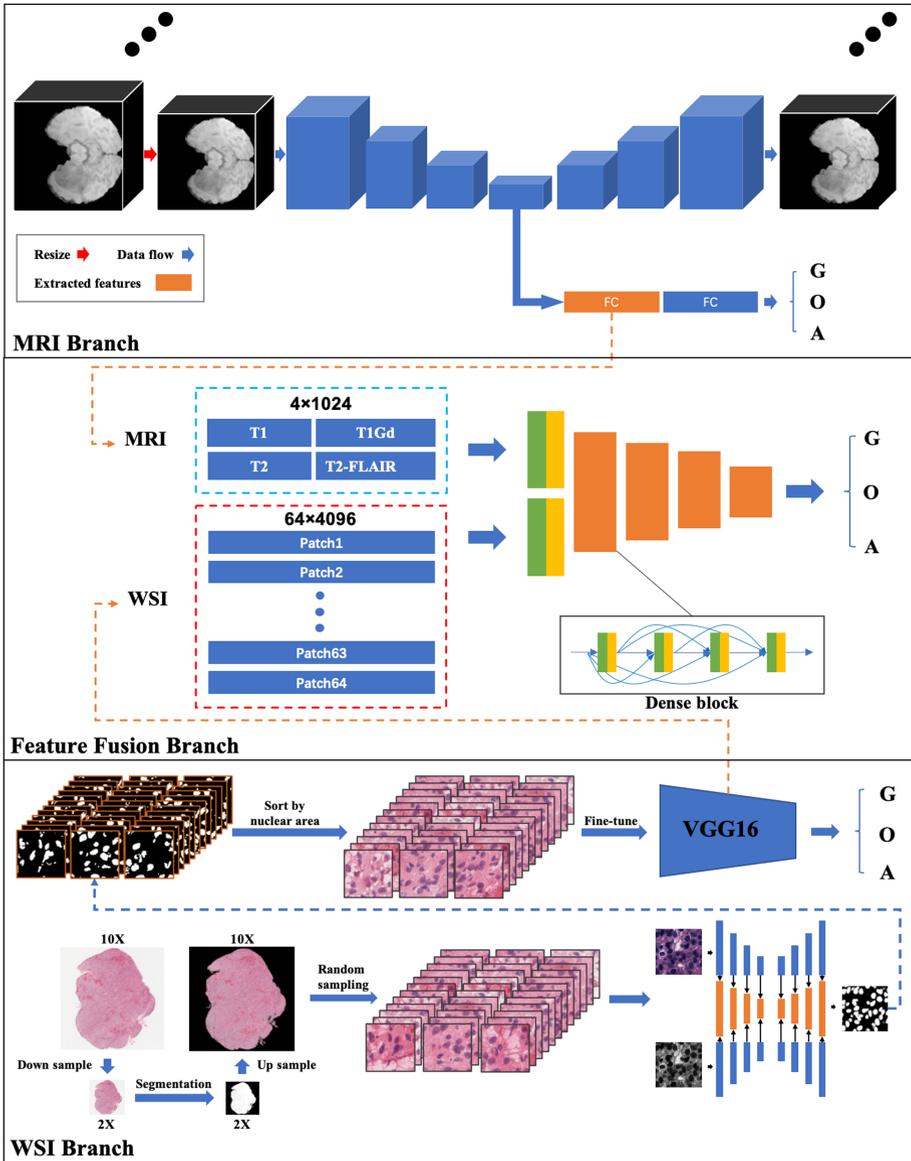


Fig. 1. Schematic diagram of the pipeline. MRI and WSI branches are proposed to extract radiology and pathology features. Feature fusion branch is proposed to aggregate multi-modal features and predict final classification results. G, O, and A denote Glioblastoma, Oligodendroglioma and Astrocytoma, respectively. The details of each branch are shown in corresponding subfigures.

network. These two feature extraction models are directly guided by the classification labels. Finally, we fused the extracted features from multi-modal images and used a densely connected neural network to predict the final classification results. Three evaluation measurements of F1-Score, Cohen’s Kappa, and Balanced Accuracy achieve 0.943, 0.903, and 0.889 on the validation set.

2 Related Works

In the CPM-RadPath 2019 MICCAI challenge [6], several works have achieved great performance on automatic brain tumor classification with multi-modal images. Pei et al. [11] segmented the brain tumor from the MRI sequence, and then classified it by a regular 3D CNN model. But they did not discover the massive information in WSI. Ma et al. [7] used two convolutional neural networks for radiology and pathology images respectively. ResNet34 and ResNet50 were directly applied to extract features from WSI grayscale patches and classified them. 3D DenseNet was employed for MRI sequence. A further regression model was introduced for the inference. Chan et al. [3] grouped WSI tiles into several clusters in an unsupervised manner and applied a random forest for final prediction. Xue et al. [15] proposed a multi-modal tumor segmentation network by leveraging the information from four MRI sequences. Then a two-branch network for both MR images and pathology images was introduced for classification.

Different from the previous works, our proposed model is segmentation-free for MRI sequences. In the meanwhile, the area where tumor cells gather is regarded as the representative region of the tumor for the feature extraction of pathology images. A deep neural network is finally applied to aggregate the multi-modal image features and to make prediction.

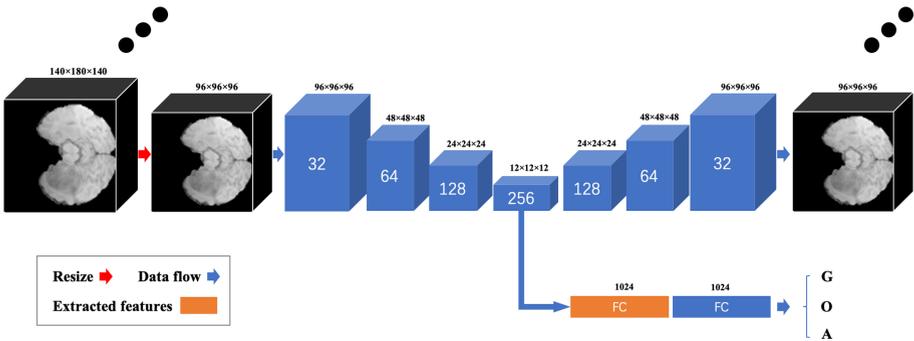


Fig. 2. Network architecture of MRI branch. It is designed as a multi-task learning model with a self-supervised feature reconstruction task and a classification task. The resolution of feature maps in each layer is shown on the top of each block. The number of the channels is shown inside the block.

3 Methods

Figure 1 demonstrates the complete classification model, which is constructed by three individual branches. The MRI branch and the WSI branch serve for radiology and pathological image feature extraction, respectively. A feature fusion branch is designed for aggregating multi-modal features and predicting the subtypes of the glioma.

3.1 Radiological Features Extraction

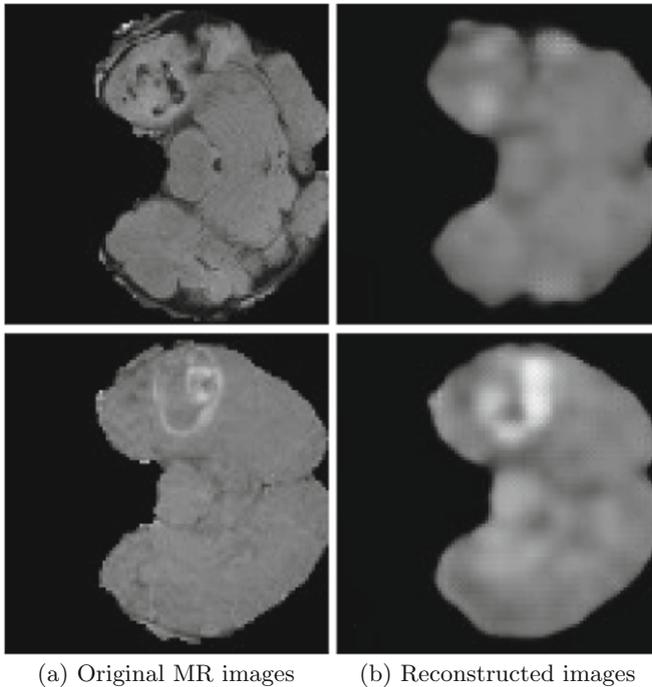


Fig. 3. The original MR images and the reconstructed images.

To extract the image features of the MRI volume, previous approaches commonly first segment the tumor lesion, and then extract the feature of the tumor. However, the accuracy of the segmentation result would directly affect the final prediction results. Moreover, since malignant tumor is invasive, the surrounding area of the tumor is also valuable for the assessment of the tumor. Therefore, we proposed a multi-task learning network to extract the features of MRI volume, which is segmentation-free. Figure 2 shows the network details. The first task is a self-supervised learning model with an autoencoder-decoder structure. It is

a common feature compression scheme by encoding the latent features of the input and reconstructing the original input. Figure 3 demonstrates the original images and the reconstructed images. However, this scheme tends to learn the pixel-wise features instead of high-level semantic features. So in the second task, we force the network to learn and predict the subtypes of the tumor from the latent features. L1 loss and cross entropy loss are used for the reconstruction and the classification tasks respectively.

$$\mathcal{L}_{L1} = \frac{1}{N} \sum_{i=1}^N |y_i - y'_i| \tag{1}$$

$$\mathcal{L}_{ce}(x) = - \sum_{i=1}^K p(x_i) \log p(x_i) \tag{2}$$

where y'_i and y_i denote the i -th pixel value of the reconstruction image and the input image respectively. N is the total number of pixels. K denote the number of tumor types.

The feature vector from the first fully connected layer of the second task are extracted as the feature representation of the MRI volume. Each MRI sequence (T1, T1Gd, T2, and T2-FLAIR) has a 1×1024 feature vector. The concatenated feature vectors of all the MRI sequence form the final feature representation of the radiological images.

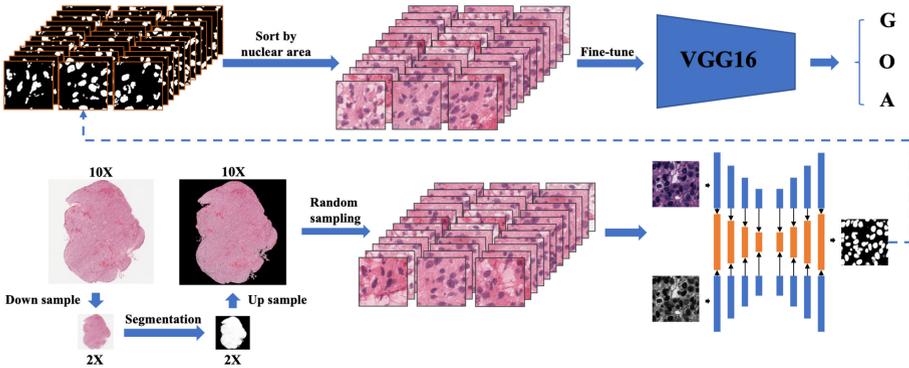


Fig. 4. Model architecture of WSI branch. The input WSI image was firstly downsampled to 2× magnification for a rough segmentation of the tissue. Then we randomly sampled a large amount of patches from the segmented region. A nuclei segmentation approach [16] was applied for each patch sample. After sorting by the area of nuclei, 64 patches with the largest nuclei area were selected and passed into a classification network (VGG). The features of WSI image can be obtained from the fully connected layer of VGG network.

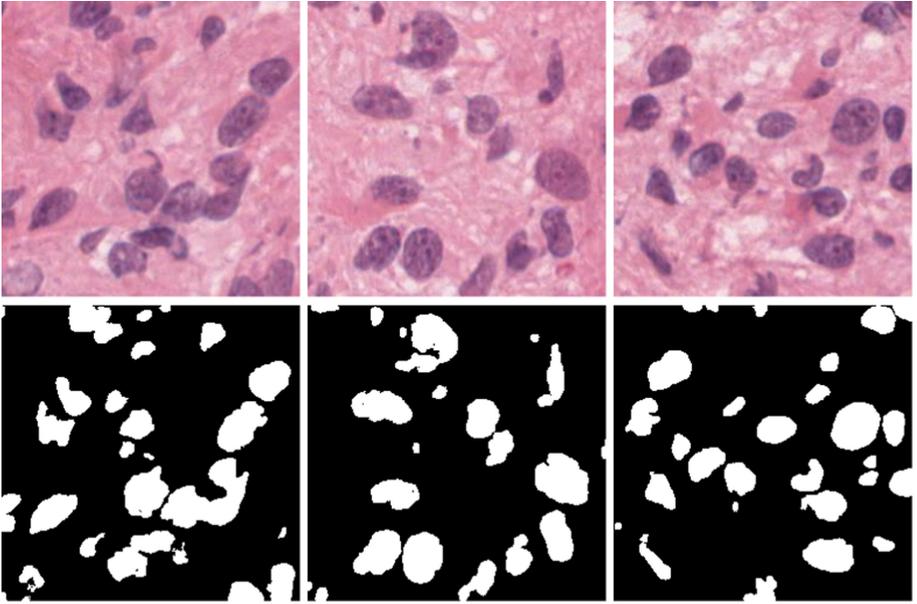


Fig. 5. Nuclei segmentation results by Triple U-net [16].

3.2 Pathological Features Extraction

Due to the gigapixel of WSI, we proposed a step-by-step procedure to quantify the pathological image, as shown in Fig. 4. We first performed a segmentation method under 2X magnification to roughly segment the tissue. Specifically, we introduced a color deconvolution algorithm [12] to map the WSI into H (Hematoxylin) channel and E (Eosin) channel, and then used the Otsu [10] to segment the foreground and background with the intersection area of the two channels as the rough segmentation mask of the tissue. Then we upsample the mask from 2X to 10X magnification to obtain the segmented region at 10X magnification. After that, we randomly sample 1000 patches (256×256) from the segmented region under 10X magnification and applied a nuclei segmentation network, Triple U-net [16]. Figure 5 shows the nuclei segmentation results. Note that, the nuclei segmentation network was trained on a public dataset MoNuSeg [5]. 64 patches with the largest nuclei area, which indicate the areas with dense tumor cells, they were selected to be the representative patches of the WSI. A VGG16 network [13] was introduced as the backbone of feature extraction and classification. The VGG16 was optimized by a cross entropy loss as Eq. 2.

3.3 Features Fusion Branch

Given the representative features extracted from the radiological and pathological images, we can get (1×4096) MRI features and (64×4096) WSI features.

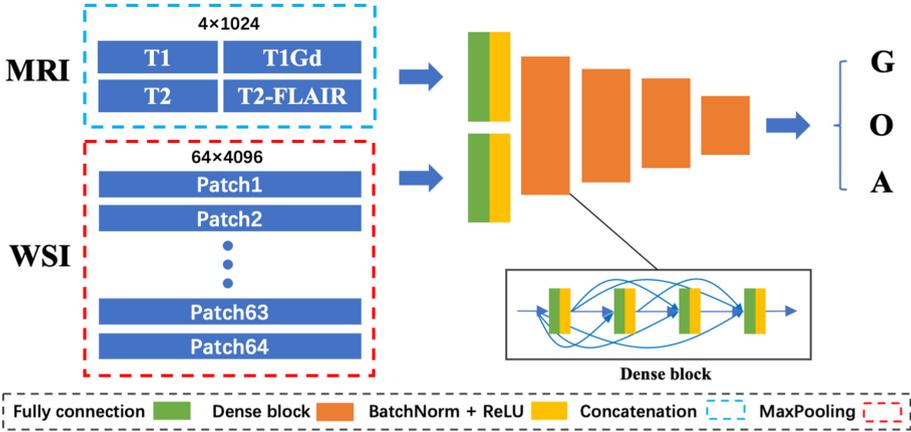


Fig. 6. Feature fusion branch. Each WSI sequence is quantified as a 1×1024 feature vector. Four MRI sequences form a 1×4096 feature vector which is the final radiological image features. We can obtain a 1×1024 feature vector of the pathological image features by maxpooling the 64 selected WSI patches features. Then the features of each modal are passed into a densely connected neural network [4] and predict the classification result.

Maxpooling operation was performed to the pathological features to downsample the features to the same dimension with the radiological features (1×4096). Then the features from two modalities were passed into the feature fusion branch, as shown in Fig. 6. Densely connected network [4] was applied to predict the subtypes of the tumor. Cross entropy loss was utilized to optimize the network.

3.4 Implementation and Training Details

All networks were implemented on Pytorch 1.5.0 and ran on a workstation equipped with an NVIDIA GeForce RTX 2080 Ti. The learning rates of the MRI branch, WSI branch, and densely connected network are 0.001, 0.0001, 0.001, respectively, and all with the learning rate decay of 0.96. All the training processes used adam optimizer without applying the dropout layer. Triple U-net [16] was trained on the public dataset from another MICCAI challenge MoNuSeg [5]. All the other models in the classification pipeline were not pre-trained.

4 Results

4.1 Quantitative Comparison

We evaluate the property of our pipeline with F1-Score, Cohen’s Kappa, and Balanced Accuracy calculated on the validation set and compare with the top four models [3, 7, 11, 15] in CPM-RadPath 2019 MICCAI challenge. Note that, all the quantitative results are from their respective papers. As is shown in Table 1.

Our pipeline achieved a classification result very close to the manual label with F1-Score of 0.943, Cohen’s Kappa of 0.903, and Balanced Accuracy of 0.889. It outperforms three existing models and get the same performance with the champion of CPM-RadPath-2019 in the validation set.

Table 1. Quantitative evaluation of classification results.

Models	F1-Score	Cohen’s Kappa	Balanced accuracy
Chan et al. [3]	–	–	0.780
Xue et al. [15]	–	–	0.849
Pei et al. [11]	0.829	0.715	0.794
Ma et al. [7]	0.943	0.903	0.889
Ours	0.943	0.903	0.889

4.2 Timing Statistics

We randomly select a patient from the test set with ID: CPM19.CBICA.ART.1 and provide comprehensive timing statistics of in Table 2. The resized resolution of MRI sequences from this patient is $96 \times 96 \times 96$. The resolution of the whole slide image is 108528×92767 . In MRI branch, feature extraction for 4 MRI sequences takes 10 s. In WSI branch, because of the huge resolution of the whole slide image, it takes 701 s to read the WSI image and 227 s to sample patches, including nuclei segmentation. The feature extraction of 64 WSI patches takes 8 s. And the final prediction only takes 2 s. Note that, the variation of the resolutions of WSI images may lead to vibration of total running time. In the future, the timing performance can be further optimized if we can effectively decrease the time of reading WSI images.

Table 2. Timing statistics of each step (second).

MRI branch	WSI branch			Prediction
	Read image	Patch sampling	Feature extraction	
10	701	227	8	2

5 Conclusion

We propose an intuitive and fully automatic pipeline for glioma classification with the input 4 MRI sequences and a H&E stained whole slide image. The proposed pipeline can effectively extract and aggregate multi-modal image features

and predict the subtypes of glioma without the necessity of any additional labels. Our model was examined on the validation set and gave a promising result.

Since our proposed pipeline has two feature extraction branches for both radiology and pathology images. And these two branches are supervised by the groundtruth labels of tumor classes. So even we omit one of the modality, the feature extraction models can still be utilized for classification. Each feature extraction branch itself can be utilized to predict the class of glioma. But lack of any modality will harm the classification performance.

Acknowledgements. This work was supported by the National Key R&D Program of China (No.2017YFC1309100), the National Science Fund for Distinguished Young Scholars (No.81925023), the National Natural Science Foundation of China (No. 81771912) and Science and Technology Planning Project of Guangdong Province (No.2017B020227012).

References

1. Annette, M.M., Jennie, W.T., John, K.W., Margaret, R.W.: Genetic and molecular epidemiology of adult diffuse glioma. *Nat. Rev. Neurol.* **15**, 405–417 (2019)
2. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., Jemal, A.: Global cancer statistics 2018: globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **68**(6), 394–424 (2018)
3. Chan, H.-W., Weng, Y.-T., Huang, T.-Y.: Automatic classification of brain tumor types with the MRI scans and histopathology images. In: Crimi, A., Bakas, S. (eds.) *BrainLes 2019*. LNCS, vol. 11993, pp. 353–359. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-46643-5_35
4. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4700–4708 (2017)
5. Kumar, N., Verma, R., Sharma, S., Bhargava, S., Vahadane, A., Sethi, A.: A dataset and a technique for generalized nuclear segmentation for computational pathology. *IEEE Trans. Med. Imaging* **36**(7), 1550–1560 (2017)
6. Kurc, T., et al.: Segmentation and classification in digital pathology for glioma research: challenges and deep learning approaches. *Front. Neurosci.* **14**, 27 (2020). <https://doi.org/10.3389/fnins.2020.00027>
7. Ma, X., Jia, F.: Brain tumor classification with multimodal MR and pathology images. In: Crimi, A., Bakas, S. (eds.) *BrainLes 2019*. LNCS, vol. 11993, pp. 343–352. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-46643-5_34
8. Omuro, A., DeAngelis, L.M.: Glioblastoma and other malignant gliomas: a clinical review. *JAMA* **310**(17), 1842–1850 (2013)
9. Ostrom, Q.T., et al.: The epidemiology of glioma in adults: a “state of the science” review. *Neuro-Oncol.* **16**(7), 896–913 (2014)
10. Otsu, N.: A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* **9**(1), 62–66 (1979)
11. Pei, L., Vidyaratne, L., Hsu, W.W., Rahman, M.M., Iftekharuddin, K.M.: Brain tumor classification using 3D convolutional neural network. In: Crimi, A., Bakas, S. (eds.) *BrainLes 2019*. LNCS, vol. 11993, pp. 335–342. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-46643-5_33

12. Ruifrok, A.C., Johnston, D.A., et al.: Quantification of histochemical staining by color deconvolution. *Anal. Quant. Cytol. Histol.* **23**(4), 291–299 (2001)
13. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *Computer Science* (2014)
14. Wesseling, P., Capper, D.: Who 2016 classification of gliomas. *Neuropathol. Appl. Neurobiol.* **44**(2), 139–150 (2018)
15. Xue, Y., et al.: Brain tumor classification with tumor segmentations and a dual path residual convolutional neural network from MRI and pathology images. In: Crimi, Alessandro, Bakas, Spyridon (eds.) *BrainLes 2019*. LNCS, vol. 11993, pp. 360–367. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-46643-5_36
16. Zhao, B., et al.: Triple u-net: hematoxylin-aware nuclei segmentation with progressive dense feature aggregation. *Med. Image Anal.* **65**, 101786 (2020)