Brain Tumor Detection and Classification from Multi-sequence MRI: Study Using ConvNets

Subhashis Banerjee[1,2]([✉]), Sushmita Mitra[1], Francesco Masulli[3], and Stefano Rovetta[3]

[1] Machine Intelligence Unit, Indian Statistical Institute, Kolkata, India
mail.sb88@gmail.com, sushmita@isical.ac.in
[2] Department of CSE, University of Calcutta, Kolkata, India
[3] DIBRIS, University of Genova, Genoa, Italy
{francesco.masulli,stefano.rovetta}@unige.it

Abstract. In this paper, we thoroughly investigate the power of Deep Convolutional Neural Networks (ConvNets) for classification of brain tumors using multi-sequence MR images. First we propose three ConvNets, which are trained from scratch, on MRI patches, slices, and multi-planar volumetric slices. The suitability of transfer learning for the task is next studied by applying two existing ConvNets models (VGGNet and ResNet) pre-trained on ImageNet dataset, through fine-tuning of the last few layers. Leave-one-patient-out (LOPO) testing scheme is used to evaluate the performance of the ConvNets. Results demonstrate that ConvNet achieves better accuracy in all cases where the model is trained on the multi-planar volumetric dataset. Unlike conventional models, it obtains a testing accuracy of 97% without any additional effort towards extraction and selection of features. We also study the properties of self-learned kernels/filters in different layers, through visualization of the intermediate layer outputs.

Keywords: Convolutional neural network · Deep learning · Brain tumor · Glioblastoma multiforme · Multi-sequence MRI · Transfer learning

1 Introduction

Glioblastoma Multiforme constitute 80% of all malignant brain tumors originating from the glial cells in the central nervous system. Based on the aggressiveness and infiltrative nature of the gliomas the World Health Organization (WHO) broadly classified them into two categories, viz. Low-grade gliomas (LGG), consisting of low-grade and intermediate-grade gliomas (WHO grades II and III), and high-grade gliomas (HGG) or glioblastoma multiforme (GBM) (WHO grade IV) [1]. Although most of the LGG tumors have slower growth rate compared to HGG and are responsive to treatment, there is a subgroup of LGG tumors

which (if not diagnosed earlier and left untreated) can lead to GBM. Histological grading, based on stereotactic biopsy test, is the gold standard for detecting the grade of brain tumors. The biopsy procedure requires the neurosurgeon to drill a small hole into the skull guided by MRI, from which a sample of the tissue is collected. There are many risk factors involving the biopsy test, including bleeding from the tumor and brain due to the biopsy needle, which can cause severe migraine, stroke, coma and even death. Other risks involve infection or seizures [2] and misleading histological grading [3]. In this context multi-sequence MRI plays a major role in the detection, diagnosis, and management of brain cancers in a non-invasive manner. Decoding of tumor phenotype using noninvasive imaging is a recent field of research, known as *Radiomics* [4], and involves the extraction of a large number of quantitative imaging features that may not be apparent to the human eye. Quantitative imaging features, extracted from MR images, have been investigated in literature for the assessment of brain tumors [5]. Ref. [6] presents an adaptive neuro-fuzzy classifier, based on linguistic hedges (ANFC-LH), for predicting the brain tumor grade using 56 3D quantitative MRI features extracted from the corresponding segmented tumor volume(s).

   Although the techniques demonstrate good disease classification, their dependence on hand-crafted features requires extensive domain knowledge, involves human bias, and is problem-specific. Subsequently manual or semi-automatic localization and segmentation of the region of interest (ROI) or volume of interest (VOI) is also needed to extract the quantitative imaging features [7]. ConvNets offer state-of-the-art framework for image recognition or classification [8]. These networks automatically learn mid-level and high-level representations or abstractions from the input training data, in the form of convolution filters that are updated during the training process. It works directly on raw input (image) data, and learn the underlying representative features of the input which are hierarchically complex, thereby ruling out the need for specialized hand-crafted image features. However training a ConvNet from scratch is generally difficult because it essentially requires large training data. In medical applications data is typically scarce, and expert annotation is expensive. Transfer learning offers a promising alternative, in case of inadequate data, to fine tune a ConvNet pre-trained on a large set of available labeled images from some other category [9].

   In this paper we exhaustively investigate the performance of ConvNets, with and without transfer learning, for non-invasive brain tumor detection and grade prediction from multi-sequence MRI. Tumors are typically heterogeneous, depending on cancer subtypes, and contain a mixture of structural and patch-level variability. Prediction of the grade of a tumor may thus be based on either the image patch containing the tumor, or the 2D MRI slice containing the image of the whole brain including the tumor, or the 3D MRI volume encompassing the full image of the head enclosing the tumor. While in the first case only the tumor patch is necessary as input, the other two cases require the ConvNet to learn to localize the ROI (or VOI) followed by its classification. Therefore, the first case needs only classification while the other two cases additionally require detection or localization. Since the performance and complexity of ConvNets

depend on the difficulty level of the problem and the type of input data representation, we prepare here three kinds viz. (i) Patch-based, (ii) Slice-based, and (iii) Volume-based data, from the original MRI dataset. Three ConvNet models are developed corresponding to each case, and trained from scratch. We also compare two state-of-the-art ConvNet architectures, viz. VGGNet [10] and ResNet [8], with parameters pre-trained on ImageNet using transfer learning (via fine-tuning).

The rest of the paper is organized as follows. Section 2 provides details about the data, its preparation in patch, slice and volumetric modes, along with some preliminaries of ConvNets and transfer learning. Section 3 introduces the proposed ConvNet architectures. Section 4 describes the experimental results, demonstrating the effectiveness in terms of both qualitative and quantitative. Finally conclusions are provided in Sect. 5.

2   Materials and Methods

In this section we provide a brief description of the data preparation at three levels of resolution, followed by an introduction to convolutional neural networks and transfer learning.

2.1   Brain Tumor Data

All experiments are performed on the TCGA-GBM [11] and TCGA-LGG [12] datasets, downloaded from The Cancer Imaging Archive (TCIA) [13]. The TCGA GBM and LGG dataset consists of 262 and 199 cases. We consider four MRI sequences for a patient MRI scan, encompassing native (T1) and post-contrast enhanced T1-weighted (T1C), T2-weighted (T2), and T2 Fluid-Attenuated Inversion Recovery (FLAIR). Since the available data is inadequate to train a 3D ConvNet model, here we formulate 2D ConvNet models based on the MRI patches (encompassing the tumor region) and slices, followed by a multi-planar slice-based ConvNet model that incorporates the volumetric information as well.

Patch-Based Dataset: The slice with the largest tumor region is first identified. Keeping this slice in the middle, a set of slices before and after it are considered for extracting 2D patches containing the tumor regions using a bounding-box. This bounding-box is marked, corresponding to each slice, based on the ground truth image. The enclosed image region is then extracted. We use a set of 20 slices for extracting the patches. In case of MRI volumes from HGG (LGG) patients, four (ten) 2D patches [with a skip over 5 (4) slices] are extracted for each of the MR sequences. Therefore a total of $262 \times 4 = 1048$ HGG and $199 \times 5 = 995$ LGG patches, with four channels each, constitute this dataset.

Slice-Based Dataset: Complete 2D slices, with visible tumor region, are extracted from the MRI volume. The slice with the largest tumor region, along with a set of 20 slices before and after it, are extracted from the MRI volume in

a sequence similar to that of the patch-based approach. While for HGG patients 4 (with a skip over 5) slices are extracted, in the case of LGG patients 10 (with a skip of 2) slices are used.

Multi-planar Volumetric Dataset: Here 2D MRI slices are extracted along all three anatomical planes, viz. axial (X-Z axes), coronal (Y-X axes), and sagittal (Y-Z axes), in a manner similar to that described above.

2.2   Convolutional Neural Networks

The fundamental constituents of a ConvNet consist of the input, convolution, activation, pooling and fully-connected layers. Some additional layers include the dropout, and batch-normalization layers.

Input Layer: This serves as the entry point of the ConvNet, accepting the raw pixel value of the input image. Here input is a 4-channel brain MRI patch/slice denoted by $I \in \mathbb{R}^{4 \times w \times h}$, where w and h represent the resolution of the image.

Convolution Layer: It is the core building block of a ConvNet. Each convolution layer is composed of a filter bank (set of convolutional filters/kernels of same width and height). A convolutional layer takes an image or feature maps as input, and performs the convolution operation between the input and each of these filters by sliding (as stride) the filter over the image to generate a set of (same as the number of filters) activation maps or the feature map.

Activation Layer: Output responses of the convolution and fully connected layers pass through some nonlinear activation function, such as a Rectified Linear Unit (ReLU), for transforming the data. ReLU, is a popular activation function for deep neural networks due to its computational efficiency and reduced likelihood of vanishing gradient.

Pooling Layer: This follows each convolution layer to typically reduce computational complexity by downsampling of the convoluted response maps. It combines spatially close, possibly redundant, features in the feature maps; thereby, making the representation more compact and invariant to small changes in an image like the insignificant details.

Fully-Connected Layer: The features learned through a series of convolutional and pooling layers are eventually fed to a fully-connected layer, typically a Multilayer Perceptron. The term "fully-connected" implies that every neuron in a layer is connected to every neuron of the following layer. The purpose of the fully-connected layer is to use these features for categorizing the input image into different classes, based on the training dataset.

Additional layers like Batch-Normalization reduce initial covariate shift. The cost function for the ConvNets is chosen as binary cross-entropy (for a two-class problem).

2.3   Transfer Learning

Typically the early layers of a ConvNet learn low-level image features, which are applicable to most vision tasks. The later layers, on the other hand, learn high-level features which are more application-specific. Therefore, shallow fine-tuning of the last few layers is usually sufficient for transfer learning. A common practice is to replace the last fully-connected layer of the pre-trained ConvNet with a new fully-connected layer, having as many neurons as the number of classes in the new target application. The rest of the weights, in the remaining layers, of the pre-trained network are retained. However, when the distance between the source and target applications is significant than one may need to induce deeper fine-tuning. This is equivalent to training a shallow neural network with one or more hidden layers. An effective strategy is to initiate fine-tuning from the last layer, and then incrementally include deeper layers in the tuning process until the desired performance is achieved.

3   ConvNets for Brain Tumor Grading

The ConvNet architectures are illustrated in Fig. 1. PatchNet is trained on the patch-based dataset, and provides the probability of a patch belong to HGG or LGG. SliceNet gets trained on the slice-based dataset, and predicts the probability of a slice being from HGG or LGG. Finally VolumeNet is trained on the multi-planar volumetric dataset, and predicts the grade of a tumor from its 3D representation using the multi-planar 3D MRI data. We use filters of size (3×3) for our ConvNet architectures. A greater number of filters, involving deeper convolution layers, allows for more feature maps to be generated. This compensates for the decrease in size of each feature map caused by "valid" convolution and pooling layers. Due to the complexity of the problem and bigger size of the input image, the SliceNet and VolumeNet architectures are deeper as compared to the PatchNet.

Pre-trained VGGNet (16 layers), and ResNet (50 layers) architectures, trained on the ImageNet dataset, are employed for transfer learning. Even though ResNet is deeper than VGGNet, the model size of ResNet is substantially smaller due to the usage of global average pooling rather than fully-connected layers. Transferring from the non-medical to the medical image domain is achieved through fine-tuning of the last convolutional block of each model, along with the fully-connected layer (top-level classifier) of each model. Fine-tuning of a trained network is achieved by retraining on the new dataset, while involving very small weight updates.

Since the base models were trained on RGB images, and accept single input with three channels, we train and test them on the slice-based dataset involving three MR sequences ($T1C$, $T2$, $FLAIR$). The $T1C$ sequence was found to perform better than $T1$, when used in conjunction with $T2$ and $FLAIR$. The following section presents the results for the proposed three level ConvNet architectures, along with that of the fine-tuned models involving transfer learning.
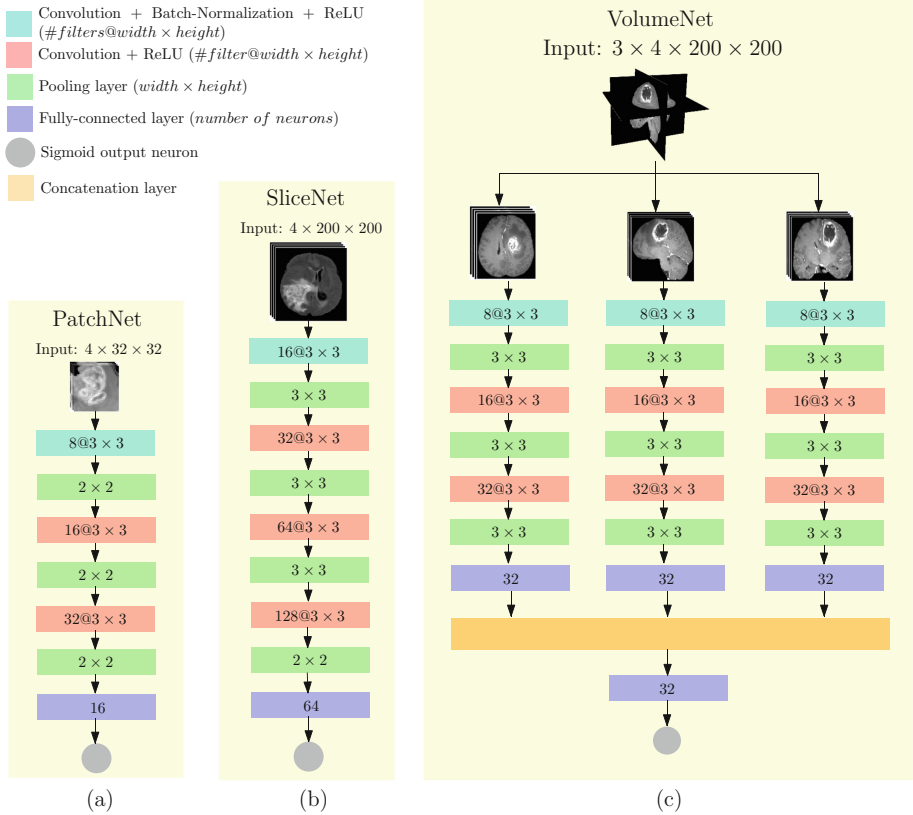
Fig. 1. Three level ConvNet architectures (a) PatchNet, (b) SliceNet, and (c) VolumeNet.

4   Experimental Results

The ConvNets were developed using TensorFlow, with Keras in Python. The experiments were performed on a desktop machine with Intel i7 CPU (clock speed 3.40 GHz), having 4 cores, 32 GB RAM, and NVIDIA GeForce GTX 1080 GPU with 8 GB VRAM. The operating system was Ubuntu 16.04. The quantitative and qualitative evaluation of the resuls are elaborated below.

We used leave-one-patient-out (LOPO) test scheme for quantitative evaluation. Although LOPO test scheme is computationally expensive, it allows availability of more data as required for ConvNets training. LOPO testing is robust and well-suited to our application, with results being generated for each individual patient. Therefore, in cases of misclassification, a patient sample may be further investigated. The ConvNet models PatchNet, SliceNet, VolumeNet, were trained on the corresponding datasets using Stochastic Gradient Descent (SGD) optimization algorithm with learning rate = 0.001 and momentum = 0.9, using mini-batches of size 32 samples generated from the corresponding training

dataset. A small part of the training set (20%) was used for validating the ConvNet model after each training epoch, for parameter selection and detection of overfitting.

Since deep ConvNets entail a large number of free trainable parameters, the effective number of training samples were artificially enhanced using real-time data augmentation – through some linear transformation. Training and validation performance of the three ConvNets were measured using *Accuracy* and F_1Score. In the presence of imbalanced data one typically prefers F_1Score over *Accuracy* because the former considers both false positives and false negatives during computation. Training and validation *Accuracy* and loss, and F_1Score on the validation dataset, are presented in Fig. 2 for the three proposed ConvNets, trained from scratch, along with that for the two pre-trained ConvNets (VGGNet, and ResNet). The plots demonstrate that VolumeNet gives the highest classification performance during training, reaching maximum accuracy on the training set (100%) and the validation set (98%) within just 20 epochs. Although the performance of PatchNet and SliceNet is quite similar on the validation set (PatchNet - 90%, SliceNet - 92%), it is observed that SliceNet achieves better accuracy (95%) on the training set (perhaps due to overfitting after 50 epochs). The performance of two the pre-trained models (VGGNet and ResNet) exhibit similar results, with both achieving around 85% accuracy on the validation set. All the networks reached a plateau after the 50th epoch. This establishes the superiority of the 3D volumetric level processing of VolumeNet.
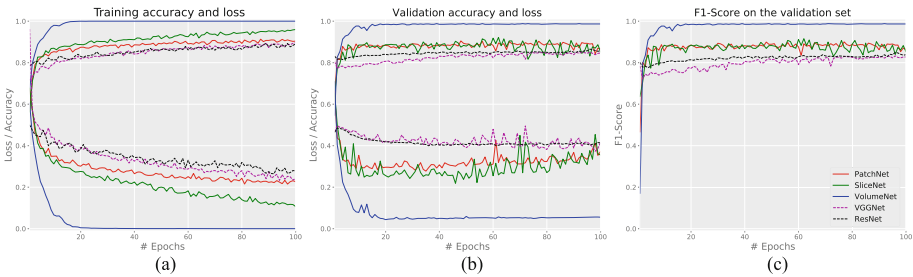

Fig. 2. Comparative performance of the networks.

After training, the networks were evaluated on the hold-out test set employing majority voting. Each patch or slice from the test dataset was from a single test patient in the LOPO framework, and was categorized as HGG or LGG. The class with maximum number of slices or patches correctly classified was indicative of the grade of the tumor. In case of equal votes the patient was marked as "ambiguous". The LOPO testing scores are displayed in Table 1. VolumeNet is observed to achieve the best LOPO test accuracy (97.19%), with zero "ambiguous" cases as compared to the other four networks. SliceNet is also found to provide good LOPO test accuracy (90.18%). Both the pre-trained models show

Table 1. Comparative LOPO test performance

ConvNets	Classified	Misclassified	Ambiguous	Accuracy
PatchNet	242	39	4	84.91%
SliceNet	257	26	2	90.18%
VolumeNet	**277**	**8**	**0**	**97.19%**
VGGNet	239	40	6	83.86%
ResNet	242	42	1	84.91%

(a) Four MRI sequences (T1, T1C, T2, FLAIR)

(c) Conv2 feature maps
Feature maps with [ ] enhancing tumor, [ ] cystic/necrotic components, [ ] edema regions highlighted

(b) Conv1 feature maps
Feature maps with [ ] tumor region highlighted

(d) Conv3 feature maps
Feature maps with [ ] different texture and shape components of the tumor
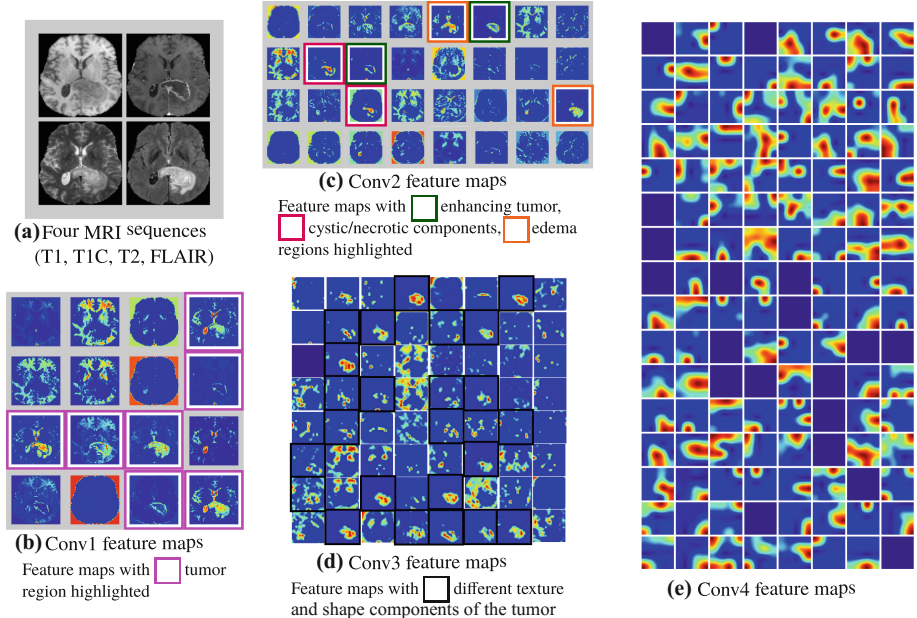
(e) Conv4 feature maps

Fig. 3. (a) Four sequences of an MRI slice from a sample HGG patient. Intermediate layer outputs/feature maps, generated by SliceNet, at different levels by (b) Conv1, (c) Conv2, (d) Conv3 and (e) Conv4.

similar LOPO test accuracy as PatchNet. This is interesting because it demonstrates that with a little fine-tuning one can achieve a test accuracy similar to that by the patch-level ConvNet trained from scratch on a specific dataset.

The ConvNets were next investigated through visual analysis of their intermediate layers. Visualizing the output of any convolution layer can help determine the description of the learned kernels. Figure 3 illustrates the intermediate convolution layer outputs (after ReLU activation) of the proposed SliceNet (Fig. 1(b)) architecture on sample MRI slices from an HGG patient.

The visualization of the first convolution layer activations (or feature maps) (Fig. 3(b)) indicates that the ConvNet has learned a variety of filters to detect

edges and distinguish between different brain tissues like white matter (WM), gray matter (GM), cerebrospinal fluid (CSF), skull and background. Most importantly, some of the filters could isolate the ROI (or the tumor); on the basis of which the whole MRI slice may be classified. Most of the feature maps generated by the second convolution layer (Fig. 3(c)) mainly highlight the tumor region and its subregions; like enhancing tumor structures, surrounding cystic/necrotic components and the edema region of the tumor. Thus the filters in the second convolution layer learn to extract deeper features from the tumor by focusing on the ROI (or tumor). The texture and shape of the tumor get enhanced in the feature maps generated from the third convolution layer (Fig. 3(d)). For example, small, distributed, irregular tumor cells get enhanced (one of the most important tumor grading criteria called "CE-Heterogeneity"). Finally the last layer (Fig. 3(e)) extracts detailed information about more discriminating features, by combining these to produce a clear distinction between images of different types of tumors.

5   Conclusion

An exhaustive study was made to demonstrate the effectiveness of ConvNets for non-invasive, automated detection and grading of brain tumors from multi-sequence MR images. Three novel ConvNet architectures were developed for distinguishing between HGG and LGG. Three level ConvNet architectures were designed to handle images at patch, slice and multi-planar modes. This was followed by exploring transfer learning for the same task, by fine-tuning two existing ConvNet models. The scheme for incorporating volumetric tumor information, using multi-planar MRI slices, achieved the best test accuracy of 97.19%. Visualization of the intermediate layer outputs/feature maps demonstrated the role of kernels/filters in the convolution layers in automatically learning to detect tumor features closely resembling different tumor grading criteria. It was also observed that existing ConvNets, trained on natural images, performed adequately just by fine-tuning their final convolution layer on the MRI dataset. This investigation allows us to conclude that deep ConvNets could be a feasible alternative to surgical biopsy for brain tumors.

References

1. Louis, D.N., et al.: The 2007 WHO classification of tumours of the central nervous system. Acta Neuropathol. **114**(2), 97–109 (2007)
2. McGirt, M.J., et al.: Independent predictors of morbidity after image-guided stereotactic brain biopsy: a risk assessment of 270 cases. J. Neurosurg. **102**(5), 897–901 (2005)
3. Chandrasoma, P.T., Smith, M.M., Apuzzo, M.L.J.: Stereotactic biopsy in the diagnosis of brain masses: comparison of results of biopsy and resected surgical specimen. Neurosurgery **24**(2), 160–165 (1989)
4. Mitra, S., Uma Shankar, B.: Medical image analysis for cancer management in natural computing framework. Inf. Sci. **306**, 111–131 (2015)

5. Zhou, M., et al.: Radiomics in brain tumor: image assessment, quantitative feature descriptors, and machine-learning approaches. Am. J. Neuroradiol. **39**(2), 208–216 (2017)
6. Banerjee, S., Mitra, S., Shankar, B.U.: Synergetic neuro-fuzzy feature selection and classification of brain tumors. In: Proceedings of IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 1–6 (2017)
7. Banerjee, S., Mitra, S., Uma Shankar, B., Hayashi, Y.: A novel GBM saliency detection model using multi-channel MRI. PLoS ONE **11**(1), e0146388 (2016)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
9. Tajbakhsh, N., et al.: Convolutional neural networks for medical image analysis: full training or fine tuning? IEEE Trans. Med. Imaging **35**(5), 1299–1312 (2016)
10. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556 (2014)
11. Scarpace, L., et al.: Radiology data from the cancer genome atlas glioblastoma multiforme [TCGA-GBM] collection. Cancer Imaging Arch. (2016)
12. Pedano, N., Flanders, A., Scarpace, L., et al.: Radiology data from the cancer genome atlas low grade glioma [TCGA-LGG] collection. Cancer Imaging Arch. (2016). https://wiki.cancerimagingarchive.net/display/Public/TCGA-LGG#aab8e738e9b547979f782c25057bfbef
13. Clark, K., et al.: The cancer imaging archive (TCIA): maintaining and operating a public information repository. J. Digit. Imaging **26**, 1045–1057 (2013)