


Article

Noninvasive Grading of Glioma Tumor Using Magnetic Resonance Imaging with Convolutional Neural Networks

Saed Khawaldeh ^{1,2,3,*}, Usama Pervaiz ^{1,2,†}, Azhar Rafiq ⁴ and Rami S. Alkhawaldeh ⁵

¹ Erasmus+ Joint Master Program in Medical Imaging and Applications, University of Girona, 17004 Girona, Spain; 12beeupervaiz@seecs.edu.pk

² Erasmus+ Joint Master Program in Medical Imaging and Applications, University of Burgundy, 21000 Dijon, France

³ Department of Electrical Engineering and Automation, Aalto University, 02150 Espoo, Finland

⁴ Department of Surgery, Virginia Commonwealth University, Richmond, VA 23298, USA; azhar.rafiq@vcuhealth.org

⁵ Department of Computer Information Systems, The University of Jordan, Aqaba 77110, Jordan; r.alkhawaldeh85@gmail.com

* Correspondence: khawaldeh.saed@gmail.com; Tel.: +34-697-637-351

† These authors contributed equally to this work and they both share the first authorship.

Received: 4 October 2017; Accepted: 18 December 2017; Published: 25 December 2017

Abstract: In recent years, Convolutional Neural Networks (ConvNets) have rapidly emerged as a widespread machine learning technique in a number of applications especially in the area of medical image classification and segmentation. In this paper, we propose a novel approach that uses ConvNet for classifying brain medical images into healthy and unhealthy brain images. The unhealthy images of brain tumors are categorized also into low grades and high grades. In particular, we use the modified version of the Alex Krizhevsky network (AlexNet) deep learning architecture on magnetic resonance images as a potential tumor classification technique. The classification is performed on the whole image where the labels in the training set are at the image level rather than the pixel level. The results showed a reasonable performance in characterizing the brain medical images with an accuracy of 91.16%.

Keywords: brain tumor classification; glioblastoma; convolutional neural network; magnetic resonance imaging

1. Introduction

Brain tumors can be cancerous or non-cancerous. In 2016, the World Health Organization (WHO) reclassified tumor types of the central nervous system into a more accurate system of brain tumor classification by integrating molecular information with the traditional histology markers [1,2]. Some of the most common brain tumors are gliomas, and they form from supportive cells in the brain called glial cells. There are different types of glial tumors such as: astrocytoma, oligodendroglioma and glioblastoma. The astrocytoma is the most common type of glioma and is formed by star-shaped cells known as astrocytes [3]. The overall classification of astrocytomas by the World health Organization is into four grades according to how abnormal the tumor cells look under the microscope and their rate of growth (I–IV) [4]. Grade I cancerous growths can be usually cured by surgical resection, as they have little proliferative capability. Grade II cancerous growths have a patient survival average of 5–15 years, as they have a relatively small proliferative potential. Grade III cancerous growths have greater malignancy, and they exhibit nuclear atypia and brisk mitotic capacity. Grade IV gliomas, which are known also as Glioblastoma Multiform (GBM), are considered as the most aggressive cancer subtype

with the presence of microvascular proliferation and pseudopalisading necrosis. More importantly, grading of brain tumors is crucial for determining the survival rate; e.g., Grade I has the highest overall survival, and Grade IV has the poorest overall survival. The grade of the glioma tumor during initial diagnosis and prognosis is essential to determine appropriate treatment options [5,6].

Typically, for the initial characterization of the tumor, Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) scans are used to produce detailed images of the brain. However, for the further classification of tumors into sub-grades, a biopsy of the tumor is necessary for detailed examination by a pathologist [7]. Tissue biopsy is an invasive procedure and is time consuming, since it requires the tissue sample to be sent to a laboratory where a pathologist conducts extensive examination and classification [8]. The technology for imaging tissue has increased in resolution to identify smaller lesions causing a greater dependency on imaging for disease diagnosis. This has resulted in the use of multiple imaging technologies for cross-referencing a suspected clinical case and achieving greater accuracy in diagnosis. The consequence of more numerous imaging studies per patient is the requirement for the integration of additional technologies to aid in the diagnosis. Technologies such as computer-assisted diagnosis methods are being developed to provide supportive diagnostic tools for analyzing medical images and identifying disease, as well as grading of brain tumors [9].

An increasing number of medical imaging techniques which align with computer-based classification and segmentation algorithms are also being examined and validated by researchers. These analytic methodologies are being applied to different types of medical images for different clinical purposes including cancer tumors' staging [10]. Such innovations also address the challenge to grading tumors by human interpretation of images from multiple readers since there is the possibility of inter-reader variability when determining the tumors' grade based on the visual features of lesions. Therefore, an automated image analytic process is sought for the classification of brain tumors that would have the capacity to quantitatively assist in more objective diagnosis [11]. The task of brain tumor image analysis is very challenging with the use of traditional machine learning algorithms for glioma cancers since there are no well-defined characteristics of the tumor, and it requires accurate differentiation of the lesion from normal tissues surrounding the cancer. The efforts to utilize image analysis for cancer diagnosis are in contrast to more recent work on genetic analysis of tissue samples [12]. Such genomic analysis requires oncogenes to be identified using functional assays from tumors and is based on the premise that cancers are a genomic disease. The Cancer Genome Atlas (TCGA) Research Network has analyzed numerous human tumors to identify molecular alterations at the DNA, RNA, protein and epigenetic levels [13]. Similar to image-guided diagnosis and tumor classification, this molecular methodology is intended to classify the tumors and guide appropriate therapies. Additionally, there are many genomic applications available in all clinical disciplines today. One such example is in work by Verhaak et al., where they classify glioblastoma subtypes with specific alterations in genetic markers such as neurofibromatosis type 1 (NF1) and platelet-derived growth factor receptor/isocitrate dehydrogenase 1 (PDGFRA/IDH1) [14]. The intent is to better understand how the genetic alterations can be linked to alternative cells of tumor origin. These cumulative data are anticipated to serve as a framework for investigation of targeted therapies to block these molecular alterations. Additional efforts by TCGA to explore GBM suggest that Methylguanyl Methyltransferase (MGMT) can shift the genetic mutation spectrum for GBM and possibly provide options for alkylating treatment. While it is beyond the scope of this report to cover all the possible methodologies being researched to better identify brain tumors, it is worth mentioning that hyperspectral imaging has also been explored as a mechanism for tumor identification. This effort was a collaborative effort that ran till 2016 to generate a large image data stack at different spectral wavelengths. The project is described as having the capacity to "discriminate between healthy and malignant tissues in real-time during surgical procedures" and thus could be classified more as an image-guided therapy modality for cancers [15].

Literature Review

In surveying the image analysis published literature, a series of research studies describe the use of imaging methodologies in conjunction with computer-aided diagnosis for the classification and segmentation of brain tumors. For example, in work by Pereira et al. the authors implemented the Convolutional Neural Network (ConvNet) as an artificial neural network for the purpose of segmenting the brain tumor based on pixel-level labeling [16]. They obtained a Dice similarity coefficient metric of 78%, 65% and 75% for the complete, core and enhancing brain tumor regions, respectively. This work is contrasted with the work proposed in the current study since their classification of brain tumor did not consider sub-grades of tumor classification using image-level labeling. Similarly, in Ertosun's study, the authors presented an automatic grading tool of glioma using deep learning [5]. However, their tool used digital pathology images, which are acquired from an invasive imaging technique. The tool proposed in the current report uses a non-invasive imaging technique, which is based on fluid-attenuated inversion recovery (FLAIR)-weighted MR images.

In studies by Tate et al. the authors used Linear Discriminant Analysis (LDA) to classify brain tumors using 1H short-echo spectra [17]. They obtained near 90% correct classification for two out of three datasets they tested with their algorithm. In additional studies by Majós et al. the authors used single voxel proton MR spectroscopy at different values of Echo Time (TE) to classify the signal into four classes including meningioma, low-grade astrocytoma, anaplastic astrocytoma and glioblastoma-metastases [18]. They obtained an accurate classification rate of 81% on the dataset of short TE values. The classification methodology addressed in the study by Ranjith et al. attempted to classify the samples into two classes: benign and malignant. In this study, the database utilized consisted of MR spectroscopy data [19]. After implementing several machine learning methods in this study, a sensitivity of 86.1% was achieved using the random forest method. The classification solutions in both these earlier works were based on the use of single voxel proton MR spectroscopy unlike the solution proposed in our work, where we used FLAIR-weighted MR images [18,19].

Additional studies published in the literature demonstrate the use of traditional machine learning methods to provide segmentation for the brain tumor, but not classification or grading. For example, the Bayesian Model used by Corso et al. was for the purpose of detecting and segmenting brain tumor from adjacent edema in multichannel MR 3D scans [20]. This Bayesian model was demonstrated to be computationally efficient with a segmentation accuracy of up to 88% compared to earlier studies analyzing MR images [21–24]. This model also had the second shortest computation time as compared to the previous studies [21–24]. Additional studies used expectation-maximization for a fully-automated tumor segmentation [25]. Expectation-maximization estimates the Probability Density Functions (PDFs) of the brain tissue classes and the intensity heterogeneity based on using T1- and T2-weighted MR images. The results of their approach were compared with a manual and semi-automatic approaches, where it gave a comparable, but less accurate performance as compared to the performance in 3D object segmentation of brain tumors [26,27].

In image analysis studies by Zacharaki the authors applied the Support Vector Machine (SVM) algorithm to classify glioma tumors [28]. They achieved a varying accuracy between 85% and 90% on the dataset they used. Both T1- and FLAIR-weighted MR images were used to assess the designed SVM classifier. The classification accuracy, sensitivity and specificity, were respectively 85%, 87% and 79% for discrimination of metastases from gliomas and 88%, 85% and 96% for discrimination of high-grade from low-grade gliomas. The study by Lawrence et al. inspired us to choose ConvNet as the deep Artificial Neural Network (ANN) for our analysis [29]. ConvNet is the most studied and validated methodology for image analysis tasks. ConvNet is well known for its special property called spatial invariance, such that the network is able to learn invariant features that make the convolution process invariant to translation, rotation and shifting. As a result, the differences between the same classed tumors due to the translation, rotation and shifting are dealt with using the spatial invariance property of ConvNet. Additionally, this makes the classification of brain tumors into different classes more feasible. In addition, the classification in this work is performed on image-level analysis.

This function enables ConvNet to predict a single label representing the class of the patient's MR image. These characteristics contrast with the segmentation option of analysis where the classification is performed at the image's pixel level. The image-level classification that is used in this work does not require advanced computational processing capacity to train the deep network and tune its weights and parameters.

The aims of this study were as follows:

- Propose a potential noninvasive replacement technique for the traditional invasive methods of grading brain tumors.
- Address the classification of brain tumor using MR images in conjunction with deep learning with artificial neural networks.
- Demonstrate a baseline application for using ConvNet in brain tumor grading and prove its efficiency for brain tumor staging.

2. Materials and Methods

2.1. Classification of Brain MR Images

In this study, an algorithm is designed using ConvNet for the aim of classifying varied brain images into categories of healthy brains, brains with low-grade tumor and brains with high-grade tumor. The presented work uses ConvNet as a classifier. This classifier is able to distinguish between the different categories based on the features that ConvNet learned automatically during the training process.

In theory, distinctions between healthy brains, brains with low-grade tumor and brains with high-grade tumor could be sorted by ConvNet because these differences are already apparent to the human eye. The main features for such distinction in classification are the existence of a necrotic core and enhancing rim around the tumor. Therefore, the target in this study is to utilize a neural network to automatically classify the MR images of the brain into three sub-classes:

1. MR brain images of healthy subjects.
2. MR brain images of glioma patients having low-grade glioma tumor.
3. MR brain images of glioma patients having high-grade glioma tumor.

Figure 1, which shows the FLAIR-weighted MR images, gives an example of how the brain of a healthy subject differs from the patient having a glioma cancer. One of the effects when glioma tumor spreads in the brain is that the distribution of fluids present in the brain is changed due to the formation of swelling, or edema around the necrotic or cystic core of the tumor, as portrayed in the figure below.



Figure 1. Magnetic Resonance Imaging (MRI) of healthy subject, low-grade and high-grade glioma tumor.

2.2. ConvNet Designed Architecture

ConvNet is considered as one of the feed-forward ANNs. It is used in this study for classifying brain tumors. The ConvNet is derived from the biological arrangement of the visual cortex of mammals where the main feature extracted as the input data to the network is the full connectedness style

of neurons in the network architecture. ConvNet has a set of unique characteristics such as the non-fully-connected network, which is different from traditional ANN. This means that the data neurons are linked with a smaller part of the prior layer. Another important feature of ConvNet is the depth of architecture where each layer looks like a hexahedron of neurons of specific dimensions regarding height, width and depth [30].

Figure 2 illustrates the design of our ConvNet, including the different layers in the ConvNet architecture. The input image passes through different sets of layers. These layers consist of the convolution layer, max-pooling layer and rectified linear unit (ReLU) layer. Moving to the network's rear part, the architecture includes a fully-connected layer and a softmax loss layer, which ensures that the output of the network represents the tissue class to which the input image belongs.

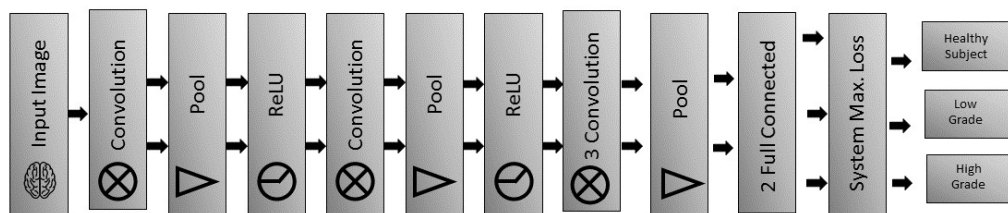


Figure 2. Layer structure of the proposed architecture.

We implemented a modified version of AlexNet where the size of the input image is 160×160 kernels [31] as shown in Figure 3. More specifically, the layers' configurations relative to the types and specifications of our network architecture are as follows, and Appendix A can be referred to for these layers' definitions.

- Convolutional layers: The filter parameters of the convolutional layer are initialized by giving random numbers from a Gaussian distribution with a spatial resolution of a 20×20 kernel array. The inputs to the convolutional layer are gray-scale images taken so that the filter depth at the input layer is 1. The number of filters is also set to 96, so that it spans the entire area of the input image.
- Pooling layers: This is a different type of pooling layer, wherein a maximum number of pooling layers shows the best performance. The size of the sliding window is set to 3×3 kernels, and the value of the stride is set to 2. The stride of 2 means the image will be re-sampled with a value of 2. The max-pooling layer, however, decreases the spatial resolution of the image equally with the stride value. The size of the sliding window and the stride value remains the same throughout different max-pooling layers in the architecture. Then, the rectified linear unit as the non-linearity layer is used. This layer converts the entire pixel of negative values to zero, which leads to making the computations simpler and also avoids further complications due to moving forward along the other layers in the ConvNet. The convolutional, max-pooling and the non-linearity layers are replicated along the architecture.
- Fully-connected layers: They consist of two layers where each one of them decreases the spatial size of the input to 1×1 . The filter depth is equal to 3, corresponding to our three classes: healthy subjects, patients having a low-grade tumor and patients having a high-grade tumor.
- Softmax loss layer: This layer is responsible for estimating the performance of the network and updating the network weights through the back-propagation process, which is based on the derivation of the loss function. This simplifies classification, because outputs are either close to 0 or 1.

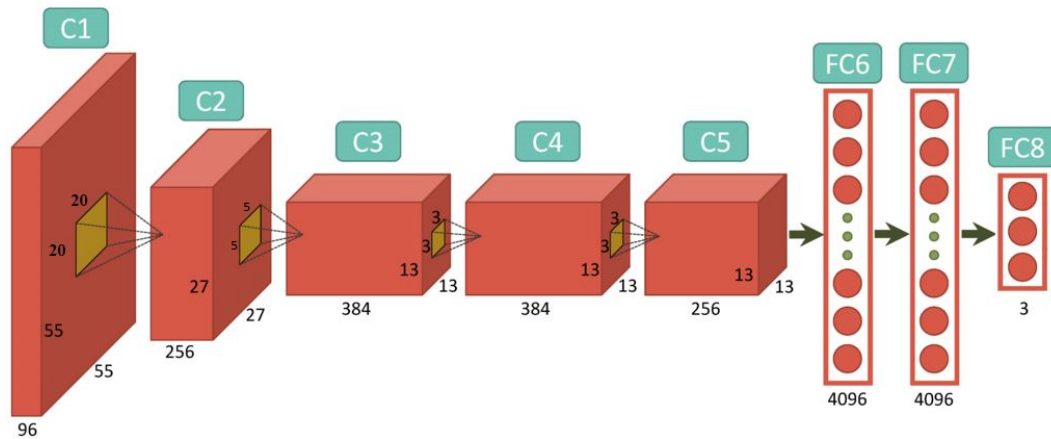


Figure 3. Simplified network architecture.

3. Implementation and Experimentation

This section describes the dataset used, how the data input is provided to the network and the preprocessing methods that are applied to the input images.

3.1. Dataset

The dataset used in this study is from the Cancer Imaging Archive (TCIA) [32,33]. The data were originally annotated and labeled by experts from Thomas Jefferson University and Henry Ford Hospitals. This dataset is composed of MR scans for 130 subjects belonging to three classes including low-grade, high-grade and healthy subjects. This public dataset is one of the most trusted online datasets; however, it has a few limitations. Firstly, the dataset is composed of MR scans from 130 subjects. However, the ground truths were only provided for 126 subjects of the total number of subjects available in the dataset. Secondly, the ground truths provided were for the entire MR scan, which implies that the information at the slice level is missing. Finally, the segmentation information was not provided, thus, the labels at the pixel level were not given. Neurologists have done a two-phase examination. The first phase consists of verifying all the annotations given by Thomas Jefferson University and Henry Ford Hospitals' experts by acting as second observers. Furthermore, they labeled the 4 unlabeled MR scans. It took approximately 2–3 min to examine each scan. This time includes the time required for loading, analyzing and providing feedback. The second phase of reviewing these scans was focused on the slice-level details. In this report, we used 2D ConvNet for the training and testing of the Computer-Aided Diagnostics (CAD) tool. Not all the slices of a full brain MRI scan of a patient with glioma tumor would show the lesion. Thus, neurologists selected only those few slices from each scan of the low-grade and high-grade tumor class that contained the lesion. The neurologists went through each scan, selecting an average of 31 slices from each MRI scan. The whole process was performed very carefully making sure that the selected slices from the low-grade and high-grade glioma class did not include any healthy slices. In this phase, It took approximately 5–7 min to analyze each scan.

In our dataset, the low-grade class consisted of Astrocytoma II and Oligodendroglioma II. The high-grade dataset class was a combination of GBM, Astrocytoma III and Oligodendroglioma III. The third and final class of the dataset was comprised of healthy subjects. These image archives from 130 subjects resulted in 4069 2D image samples in total. On average, 31 2D slices were selected out of each brain MR scan. The entire data break down of low-grade and high-grade gliomas is shown in the Table 1. Additionally, the overall data segregation with respect to gender, age and race among low-grade and high-grade glioma patients is shown in Table 2.

It is very important to diagnose glioma in the early stage for the greater prognosis of patients. However, one of the limitations of existing CAD schemes is that they perform well only on the detection

of large-sized tumors. This reemphasizes the fact that it is significantly important to take into account the size and the shape of tumors.

Table 1. Low-grade glioma and high-grade glioma composition.

Tumor Subtype	Low-Grade Glioma	High-Grade Glioma
Oligodendroglioma II	11	-
Oligodendroglioma III	-	7
Glioblastoma Multiform (GBM) IV	-	43
Astrocytoma II	30	-
Astrocytoma III	-	17

Table 2. Statistical analysis.

Variable	Low-Grade Glioma	High-Grade Glioma
Gender	Number of Cases	Number of Cases
Male	18	25
Female	16	20
Unknown	7	23
Race	Number of Cases	Number of Cases
White	29	40
Black	4	2
Asian	0	1
Unknown	8	25
Age	Number of Cases	Number of Cases
10–19	2	1
20–29	5	2
30–39	14	8
40–49	5	10
50–59	6	11
60–69	3	6
70–79	1	7
80–89	1	0
Unknown	4	23

In Table 3, the range of tumor size measurements within this patient group is displayed. The lesion size was calculated by using the largest perpendicular (x-y) cross-sectional diameter of signal abnormality (longest dimension \times perpendicular dimension) measured on a single axial image only. The 1D measurements were provided by radiologists for 99 subjects. The 2D size of the lesion area is a computational approximation by multiplying the 1D cross-sectional diameters.

Table 3. Lesion size summary.

Lesion Size	1D-x (cm)	1D-y (cm)	2D (cm ²)
Minimum	3	1.5	4.5
Mean	5.77	4.19	26.214
Maximum	>8	>8	>64

Figure 4 shows the size of lesion for each of the individual patients along both major perpendicular axes. Some of the patients in this study have a lesion size less than 2 cm. All images available from the database are used for training and evaluation of the designed classifier irrespective of the lesion sizes. Additional details regarding the detailed distribution of lesion size among the 18 classes are provided in Appendix B.

In our dataset, we have patients diagnosed with multiple gliomas, which represent approximately 2–20% of the high-grade glioma class [34]. The multiple glioma patients can be categorized into either of the following three classes: multifocal, multicentric and gliomatosis. These are further described in Appendix C.

In our report, the number of patients that had multifocal, multicentric or gliomatosis was 7, 4 or 4 patients, respectively. Additionally, in the FLAIR MR scans that we had in our dataset, various edema portions were present in both training and testing subsets. For example, our dataset had 22, 25, 16 and 36 patients who had 34–67%, 6–33%, <5% and 0% proportions of edema, respectively. Furthermore, the edema surrounding the tumor was crossing the mid-line between the white matter and the grey matter in 36 patients. The training and the testing subsets included patients from the various types of tumor presentations, which were mentioned above; this ensures a robust performance for the classifier being designed.

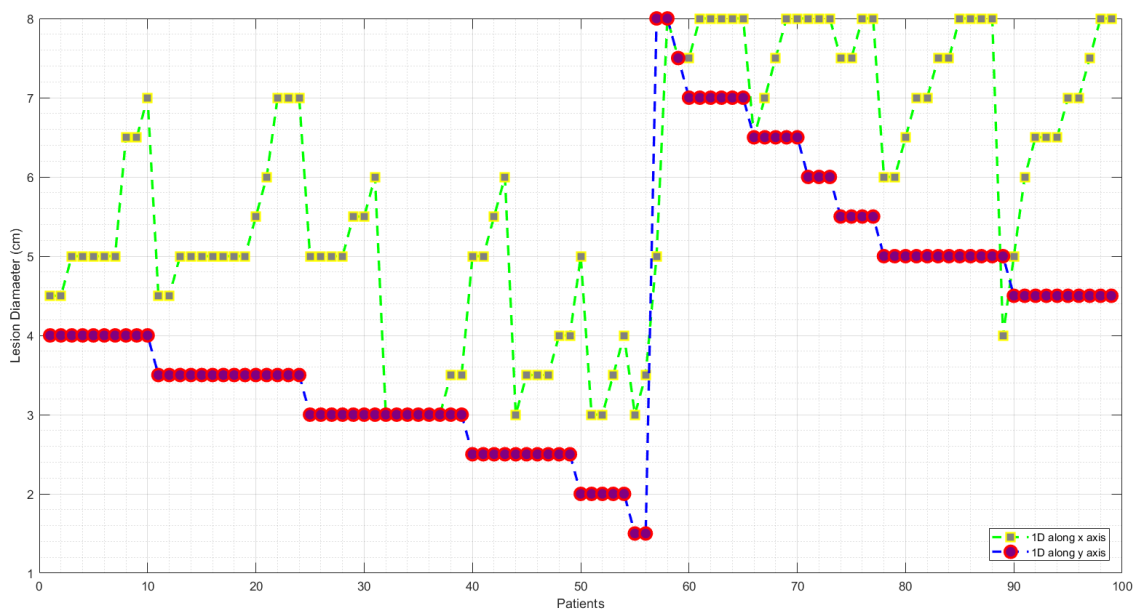


Figure 4. Lesion size (x-y). It is to be noted that we do not have the exact measurement of the size for a tumor length greater than 8 cm, so they are thresholded to an 8-cm value in the figure for simplicity.

Furthermore, we labeled the data as 3 classes: Class 1 belongs to the MR images of healthy patients; Class 2 belongs to the MR images of patients with low-grade tumor; and Class 3 belongs to the MR images of patients with high-grade tumor. Finally, we randomly divided the samples in the dataset into training, validation and testing subsets, as is shown in Table 4.

Table 4. Data division in the dataset.

Data Division	Healthy	Low-Grade	High-Grade	Total
Training	700	1031	1077	2808
Validation	155	267	252	674
Testing	133	230	224	587
Total	988	1528	1553	4069

3.2. Implementation Platform

The platform used to implement our ConvNet was Deeplearning4j (DL4J). It is considered to be the first open-source and distributed deep-learning library. Moreover, DL4J is designed to be used on

distributed GPUs and CPUs, which gives more flexibility and wider options during the implementation and experimentation. While DL4J is written in Java, it is compatible with some Java Virtual Machine (JVM) languages like Scala or Clojure. Additionally, the platform’s underlying computations are written in C, C++ and CUDA.

3.3. Pre-Processing

The input size of each image in our dataset is 256×256 . Before feeding any image to the system, the contrast of the input image is improved and normalized using the mean and Standard Deviation (STD) as suggested by Coates [35]. Specifically, *ContrastNormalization* and *WhitenData* algorithms were applied on the input images before feeding them to the ConvNet. In addition, images were down-sampled to the dimension of 160×160 as shown in Figure 5.

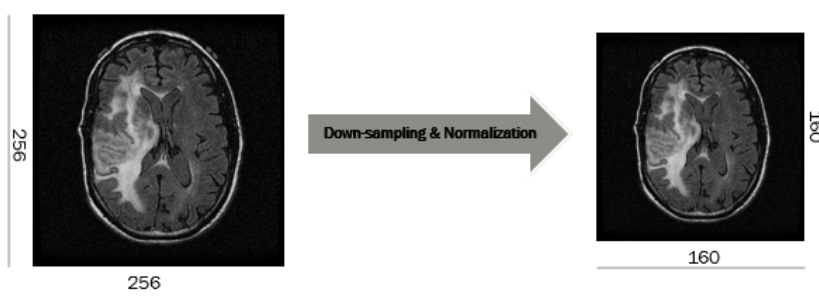


Figure 5. Down-sampling and normalization of MR images.

3.4. Parameters’ Selection

It is not intuitive what the most efficient network parameters would be for using ConvNet. It requires extensive testing and artistry to design the optimum layers’ structure. There is no pre-defined solution to select ideal values of parameters in different layers. Thus, multiple combinations of settings were attempted with different parameters for tuning and adjusting the settings according to each dataset.

Table 5 lists the different layers and their parameters, the corresponding filters and their sizes, as well as the division of data for our proposed architecture.

Table 5. The parameter design of our Convolutional Neural Network (ConvNet).

Layer	1	2	3	4	5	6	7	8	9	10	11	12	13
Type	conv	ReLU	mpool	conv	ReLU	mpool	conv	conv	conv	mpool	fc	fc	softmax
Support	20	1	3	5	1	3	3	3	3	3	1	1	1
Filter dim	1	n/a	n/a	48	n/a	n/a	192	192	128	n/a	n/a	n/a	n/a
No. of filters	96	n/a	n/a	256	n/a	n/a	384	384	256	n/a	4096	4096	3
Stride	1	n/a	3	1	n/a	3	1	1	1	3	1	1	1
Data Size	160	55	55	27	27	27	13	13	13	13	1	1	1
Data Depth	48	48	48	128	128	128	192	192	128	128	4096	4096	1

The output of ConvNet is highly sensitive to fine tuning of the parameters. The parameters like batch size, number of iterations, splitting of data and filter size dramatically affect the final results. A brief explanation of a few of the parameters is provided below:

- Epochs: The epoch refers to passing of an entire set of data through the complete architecture. At each iteration, there is random shuffling of training data. When it classifies the image wrongly, it is counted as a loss, and through back-propagation, the weights of different layers are updated, which eventually decreases the loss and error rate. The epochs are set to 40 in our case.
- Learning rate: This is considered to be a very significant and critical parameter of the network and is modified whenever the input or layer structure is modified. The learning rate is used to

determine the size of the gradient taken by the network weights. If your learning rate is small, it means you need more epochs to acquire the best trained model. In our case, the learning rate was fixed to 0.01 since this value was optimal for our dataset.

- **Weight decay:** This is a parameter required for the weight update rule, which is useful when there is no scheduled update, so it causes the weight to decay exponentially. The weight decay value is set to 0.0005 in our proposed network.
- **Batch size:** This characterizes the number of samples that will be propagated along the ConvNet. In our case, the batch size was fixed to 50.
- **Dropout:** This is a regularization technique where randomly-selected neurons drop out during training. It results in multiple independent internal representations being learned by the network. Hence, the network prevents over-fitting on the training data and is capable of improved generalization. In the proposed network, the dropout layer is added after each of the fully-connected layers with the dropout rate set to 25%. This rate means 1 in 4 inputs will be randomly excluded from each update cycle.

3.5. Model Specification

The following criteria explain the specification of our model and also evaluate its performance.

1. The total time consumed by the computer for both training and validation in our experiment was 3 h and 21 min.
2. The memory capacity required for the parameters was nearly 200 MB. The data memory used was near 400 MB. However, this can vary and depends mainly on the batch size.

Figure 6 illustrates the full details of the designed architecture, which is inspired by Krizhevsky's work [31].

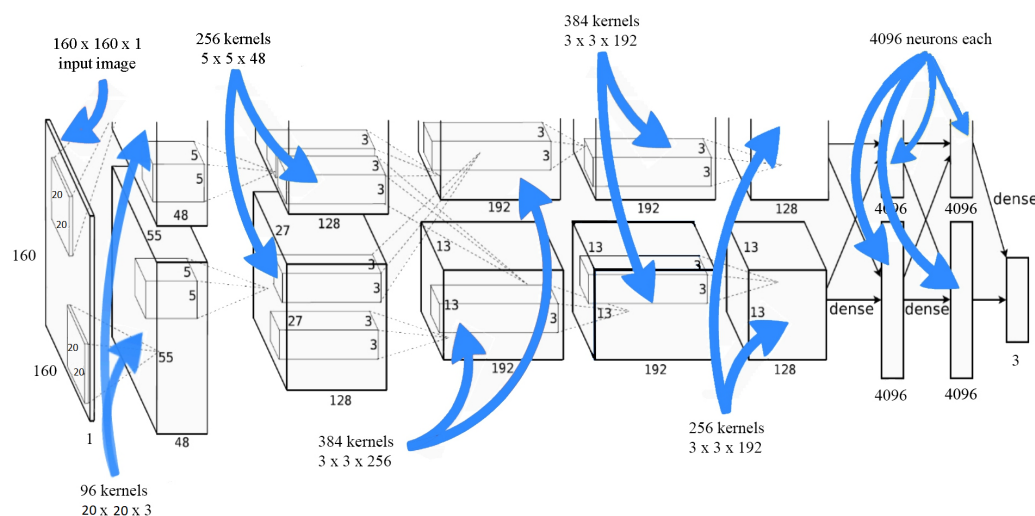


Figure 6. Detailed architecture.

4. Results and Discussion

The customized ConvNet image identification and classification system presented in this report demonstrated very accurate results in classifying input images into three main diagnostic categories. The three classification labels corresponding to the brain's state were healthy, low-grade tumor or high-grade tumor. In the evaluation phase, we used four measures to prove the efficiency of the proposed approach including accuracy, precision, recall and F1 score, which are defined in Appendix D [36]. Figure 7 illustrates the confusion matrix along with the labels of the three classes in our dataset. These four measures were calculated using the confusion matrix, which was generated after running our trained ConvNet on the testing subset consisting of 587 image samples.

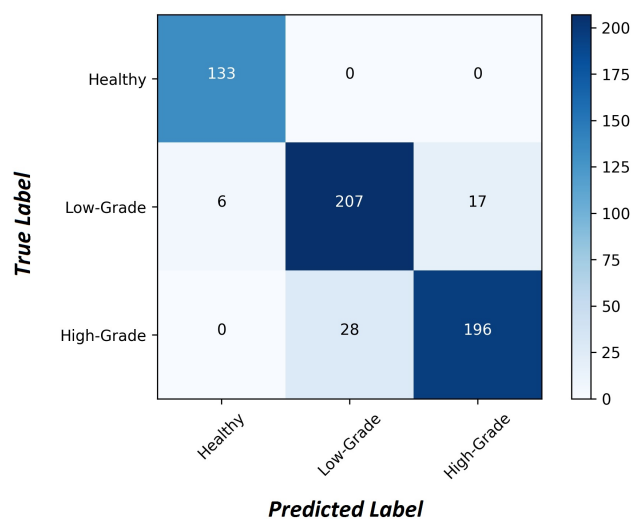


Figure 7. Confusion matrix with labels.

Table 6 lists the outcomes from measures of 587 cases using our neural network model with an accuracy of 91.3%. The accuracy metric sometimes does not effectively reflect the performance of models due to the availability of imbalance labels in the dataset or increasing number of labels [37]. Thus, three other measures are used to address such possible constraints. Precision and recall metrics represent the cases that are predicted correctly over all positive predictions and observations respectively. The precision value of 91.79% in this study indicates that our classification model accurately predicts the required label 91.79% of the time. Similarly, the required label is predicted with 92.25% in all the presented image cases. This implies better performance of the proposed architecture in classifying the brain tumor into one of the three grade levels. Moreover, we utilized the F1 score measure, which combines the precision and recall metrics to evaluate the test’s accuracy and is defined as the weighted harmonic mean of the precision and recall of the test. The F1 score balanced the two metric values and provided performance with 92.05% of predicted labels.

Table 6. Classification results.

Accuracy	0.9116
Precision	0.9179
Recall	0.9225
F1 Score	0.9205

We have selected ConvNets over other deep network architectures including recurrent or recursive neural networks (defined in Appendix E) mainly because ConvNets can classify data in a hierarchical way. This means that lower layers are used to learn low-level features like edges, blobs and corners, while the higher layers combine these basic features to identify high-level shapes. On the other hand, recurrent neural networks are suitable for the sequential data; an image would be a sequence with only one object, which does not make recurrent networks useful for this kind of classification. A recursive neural network can be seen as a generalization of the recurrent neural network, which has a specific type of skewed tree structure and is mostly applied in learning sequence and tree structures in natural language processing. Therefore, choosing ConvNets is the most optimum among all the other neural network architectures.

There can be an argument in favor of using an architecture with a lesser number of layers instead of using the modified AlexNet (AlexNet_m). To be certain, we performed a comparative analysis between the AlexNet_m and the three convolutional layer neural network (ConvNet-3). The detailed architecture

of ConvNet-3 can be seen in the table in Appendix F. With the use of ConvNet-3, we achieved an accuracy, precision, recall and F1 score of 85.71%, 85.26%, 87.68% and 86.45%, respectively. These results demonstrate that the deeper ConvNet (AlexNet_m) outperforms the less deep ConvNet (ConvNet-3), which confidently justifies our choice of a deeper network.

We have used only FLAIR MR images in this study because we wanted our system to be able to solve the classification problem achieving the highest possible performance relying only on one type of MR image. This approach shows that the proposed method is capable of achieving highly accurate results while using only one imaging modality, which is FLAIR MR images. The results from this study support a potential future practice of using less image modalities for brain tumor grading. In addition, it implies that the image acquisition time can be reduced since one imaging modality is enough to do the diagnostic classification. In [38], the authors also used only FLAIR MR images for detecting and segmenting brain tumors. The results show that with the use of only one imaging modality, FLAIR MR, they were able to obtain accurate performance with a classification precision of 87.86% and a sensitivity of 89.48%.

Another concern is related to the use of only axial MR images. We opted for the axial plane MR images as they have less noise and higher resolution in comparison to sagittal and coronal plane slices. In contrast, using blurry thin coronal and sagittal images can decrease the performance of the predictive system. Finally, it can be seen from studies in the literature such as work by Shah et al. that using axial plane images for tumor detection and classification is a common practice since it yields better overall results [39].

The 12-layer ConvNet model presented in this paper is comprised of convolutional, sub-sampling, dense and fully-connected output layers. The overall accuracy achieved is 91.16% in this study. This outcome exceeds findings from reports in the literature where the maximum accuracy for this classification problem achieved was between 85–90% [28]. In contrast, the results by Zacharaki et al. discuss binary classification with discrimination of metastases from gliomas and the discrimination of high-grade from low-grade gliomas [28]. This is differentiated from the current report, which presents results towards multiple classes of classification in order to discriminate between healthy brains, brains with low-grade glioma and brains with high-grade glioma. Additionally, the work reported here utilized only FLAIR weighted MR scans, and we did not assist the algorithm with any other MR image types such as T1- or T2-weighted images. The studies by Zacharaki et al. used both T1- and FLAIR-weighted MR images [28]. Overall, the results reported in this study are innovative and can be considered as the state of the art for classifying brain scans using ConvNet for the neural network architecture.

5. Conclusions

This paper presents a unique mechanism for configuring artificial neural network capacities to accurately classify images of brain tumors. The possible future work envisioned is to include T1-weighted and T2-weighted MR images, since this study only includes axial FLAIR-weighted MR images. Feeding 3D MR images into the system would further improve the network's performance since it will enable processing the tumor's 3D voxels beyond the validated processing of 2D slice images. However, this approach can create the limitation for the study of having a limited number of samples in the dataset, which is a challenge for complex machine learning algorithms such as ConvNet. We are confident that inclusion of 3D voxels with T1-weighted and T2-weighted MR images would help further improve the accuracy of neural networks since both tumor and edema appear to be similarly dark in T1-weighted MR images. It should also be noted that both edema and tumor look bright on T2-weighted MR images.

An additional approach to enhance the current study is increasing the number of classes that can be classified by including more sub-grades of glioma tumors. This approach would provide an additional diagnostic support to doctors and healthcare systems faced with limited staff and resources. Utilization of neural network software that incorporates using pre-trained models eliminates the need

for its users to go through the lengthy training and optimization process. The resulting technology would assist doctors to conduct large-scale classification of glioma tumor grades using non-invasive imaging techniques such as MRI in conjunction with an innovative machine learning methodology such as ConvNet. This could result in the diagnostic phase of the tumor grading process having lower costs, being less painful for patients and less time-consuming.

Acknowledgments: No funding resources are reported.

Author Contributions: Saed Khawaldeh and Usama Pervaiz designed and supervised the study. Saed Khawaldeh and Usama Pervaiz collected and prepared the dataset. Saed Khawaldeh and Usama Pervaiz programmed the model. Saed Khawaldeh and Usama Pervaiz performed the experiments. Saed Khawaldeh, Usama Pervaiz, Azhar Rafiq and Rami S. Alkhalwaldeh analyzed the results. Saed Khawaldeh and Usama Pervaiz wrote the manuscript. Saed Khawaldeh, Usama Pervaiz, Azhar Rafiq and Rami S. Alkhalwaldeh revised the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest: None of the authors of this paper have a financial or personal relationship with other people or organizations that could influence or bias the content of the paper.

Appendix A. ConvNet Types of Layers

In the standard ConvNet architecture, there are five different types of layers:

- The convolutional layer compresses a large amount of data into a smaller set of features and is the core component in ConvNets. This layer is responsible for calculating neurons' output that are locally connected to the input by performing a sliding dot product called convolution between weights and input values.
- The pooling layer reduces the spatial size of the representation to decrease the amount of parameters and computation in the network. In addition, it controls the over-fitting, and it makes data invariant to small translational changes. This layer, usually, takes the average or maximum value across disjoint patches.
- The fully-connected layer is invariably positioned as the last part of the ConvNet architecture and is responsible for assigning class scores in supervised settings.
- The ReLU layer converts any negative values coming out of the max-pooling layer to zero. Thus, after passing through the ReLU layer, there will be no negative value in the image.
- The softmax loss layer is used for the performance evaluation for each input. In a broader context, it shows the difference between the final activation layer and the ground truth.

Appendix B. Lesion Size Label Details

Table A1. Lesion size summary.

Class	Lesion Size (cm)
1	<0.5
2	0.5
3	1.0
4	1.5
5	2.0
6	2.5
7	3.0
8	3.5
9	4.0
10	4.5

Table A1. Cont.

Class	Lesion Size (cm)
11	5.0
12	5.5
13	6.0
14	6.5
15	7.0
16	7.5
17	8.0
18	>8.0

Appendix C. Multiple Gliomas

- Multifocal is defined as having at least one region of tumor, either enhancing or non-enhancing, which is not contiguous with the dominant lesion and is outside the region of signal abnormality (edema) surrounding the dominant mass. This tumor formation can be defined as resulting from dissemination or growth by an established route, spread via commissural or other pathways, or via cerebrospinal fluid (CSF) channels or local metastases.
- Multicentric are widely separated lesions in different lobes or different hemispheres of the brain that cannot be attributed to one of the previously mentioned pathways.
- Gliomatosis refers to generalized neoplastic transformation of the white matter within most of a hemisphere.

Appendix D. Performance Measures

A brief explanation of the four measures is provided below [36]:

- Accuracy refers to the closeness of a measured value to a standard or known value. In other words, it is the number of true predictions made divided by the total number of predictions made.

$$Accuracy = \frac{TP + FN}{TP + FN + TP + TN} = \frac{\text{correct predictions}}{\text{all predictions}} \quad (A1)$$

- Precision refers to the closeness of two or more measurements to each other. In other words, it is the fraction of relevant instances among the retrieved instances. However, precision is independent of accuracy since a network can be very precise, but inaccurate.

$$Precision = \frac{TP}{TP + FP} = \frac{\text{positive predicted correctly}}{\text{all positive predictions}} \quad (A2)$$

- Recall (known also as sensitivity) is the fraction of relevant instances that have been retrieved over the total relevant instances.

$$Recall = TPR = \frac{TP}{TP + FN} = \frac{TP}{P} = \frac{\text{predicted to be positive}}{\text{all positive observations}} \quad (A3)$$

- The F1 score is the weighted average of precision and recall. In other words, it is the harmonic mean of precision and recall. F1 score is an 'average' of both precision and recall. We use the harmonic mean because it is the appropriate way to average ratios (while the arithmetic mean is appropriate when it conceptually makes sense to add things up).

$$F1 = 2 \frac{Precision \times Recall}{Precision + Recall} \quad (A4)$$

Appendix E. Types of Artificial Neural Networks

In this section, various architecture types of Artificial Neural Networks (ANN) are defined. The term architecture refers to the way in which neurons of the network are connected.

- Feed-forward neural network: This is the most common type of ANN and is used in many practical applications. In this type of network, the information comes in the neural unit and flows in one direction through hidden layers until it reaches the output unit. The first layer is considered as the input, and the last layer represents the output of the network. If there is more than one hidden layer, it is called a deep neural network (DNN). DNN is a type of feed-forward ANN that has multiple hidden layers of units between the input and the output layers. They compute a series of transformations between the input and output, so that at each layer, you get a new representation of the input where things that were similar in the previous layers may become less similar or things that were dissimilar in previous layers may become more similar.
- Recurrent neural network: Another well-known architecture is called the recurrent neural network. In this type of neural network, the information can flow around in a circle. Generally, recurrent neural networks are more powerful than feed-forward networks. They have directed cycles in their connection graph, which makes it possible to get back to the start by following the arrows. These networks can remember information for a long time; furthermore, they can exhibit all sort of interesting oscillations, but they are much more difficult to train because of their complex dynamics. Recurrent, unlike feed-forward, neural networks can use their internal memory to process arbitrary sequences of inputs. Recurrent neural networks use time series information and are ideal for text and speech analysis.
- Recursive neural networks: A recursive neural network architecture is composed of a shared-weight matrix and a binary tree structure that allows the recursive network to learn varying sequences of words or parts of an image. Recursive neural networks can recover both granular structure and higher-level hierarchical structure in datasets such as images or sentences. These networks are mostly used in image scene decomposition and audio-to-text transcription.

Appendix F. Comparing Network Architectures

Table A2. Comparing network architectures: filter number \times filter size (e.g., 96×20^2), filter stride (e.g., str 2), pooling window size (e.g., pool 3^2) and the output feature map size (e.g., map size 55×55).

Model	<i>conv</i> ₁	<i>conv</i> ₂	<i>conv</i> ₃	<i>conv</i> ₄	<i>conv</i> ₅
AlexNet Modified	96×20^2 , str 2 pool 3^2 , str 2 map size 55×55	256×5^2 , str 2 pool 3^2 , str 2 27×27	384×3^2 , str 2 13×13	384×3^2 , str 2 13×13	256×3^2 , str 2 13×13
ConvNet-3 Modified	32×10^2 , str 2 pool 3^2 , str 2 map size 55×55	64×5^2 , str 2 pool 3^2 , str 2 27×27	128×3^2 , str 2 13×13	-	-

References

1. Rogers, L.; Alsrouji, O.; Wolansky, L.; Badve, C.; Tatsuoka, C.; Clancy, K. Quantitative MRI morphologic characteristics and quantitative histologic profiles in surgically proven radiation necrosis versus recurrent brain tumor (P4. 252). *Neurology* **2016**, *86* (Suppl. 16), P4-252.
2. Louis, D.N.; Perry, A.; Reifenberger, G.; Von Deimling, A.; Figarella-Branger, D.; Cavenee, W.K.; Ohgaki, H.; Wiestler, O.D.; Kleihues, P.; Ellison, D.W. The 2016 World Health Organization classification of tumors of the central nervous system: A summary. *Acta Neuropathol.* **2016**, *131*, 803–820.
3. Liffers, K.; Kolbe, K.; Westphal, M.; Lamszus, K.; Schulte, A. 264: Histone deacetylase inhibitors resensitize glioblastoma cells to EGFR-directed therapy with tyrosine kinase inhibitors after primary treatment failure. *Eur. J. Cancer* **2014**, *50*, S62.

4. Louis, D.N.; Ohgaki, H.; Wiestler, O.D.; Cavenee, W.K.; Burger, P.C.; Jouvet, A.; Scheithauer, B.W.; Kleihues, P. The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol.* **2007**, *114*, 97–109.
5. Ertosun, M.G.; Daniel, L.R. Automated grading of gliomas using deep learning in digital pathology images: A modular approach with ensemble of convolutional neural networks. *AMIA Annu. Symp. Proc.* **2015**, 2015, 1899–1908.
6. Kazantsev, A.G.; Thompson, L.M. Therapeutic application of histone deacetylase inhibitors for central nervous system disorders. *Nat. Rev. Drug Discov.* **2008**, *7*, 854–868.
7. Miller, K.D.; Siegel, R.L.; Lin, C.C.; Mariotto, A.B.; Kramer, J.L.; Rowland, J.H.; Jemal, A. Cancer treatment and survivorship statistics, 2016. *CA Cancer J. Clin.* **2016**, *66*, 271–289.
8. Bowden, S.G.; Neira, J.A.; Gill, B.J.; Ung, T.H.; Englander, Z.K.; Zanazzi, G.; Chang, P.D.; Samanamud, J.; Grinband, J.; McKhann, G.M. Sodium fluorescein facilitates guided sampling of diagnostic tumor tissue in nonenhancing gliomas. *Neurosurgery* **2017**, *13*, 307.
9. Smith-Bindman, R.; Miglioretti, D.L.; Johnson, E.; Lee, C.; Feigelson, H.S.; Flynn, M.; Greenlee, R.T.; Kruger, R.L.; Hornbrook, M.C.; Roblin Solberg, L.I.; et al. Use of diagnostic imaging studies and associated radiation exposure for patients enrolled in large integrated health care systems, 1996–2010. *JAMA* **2012**, *307*, 2400–2409.
10. Norouzi, A.; Rahim, M.S.M.; Altameem, A.; Saba, T.; Rad, A.E.; Rehman, A.; Uddin, M. Medical image segmentation methods, algorithms, and applications. *IETE Tech. Rev.* **2014**, *31*, 199–213.
11. Koob, M.; Girard, N.; Ghattas, B.; Fellah, S.; Confort-Gouny, S.; Figarella-Branger, D.; Scavarda, D. The diagnostic accuracy of multiparametric MRI to determine pediatric brain tumor grades and types. *J. Neurooncol.* **2016**, *137*, 345–353.
12. Khawaldeh, S.; Pervaiz, U.; Elsharnoby, M.; Alchalabi, A.E.; Al-Zubi, N. Taxonomic classification for living organisms using convolutional neural networks. *Genes* **2017**, *8*, 326.
13. Weinstein, J.N.; Collisson, E.A.; Mills, G.B.; Shaw, K.R.M.; Ozenberger, B.A.; Ellrott, K. The cancer genome atlas pan-cancer analysis project. *Nat. Genet.* **2013**, *45*, 1113–1120.
14. Verhaak, R.G.; Hoadley, K.A.; Purdom, E.; Wang, V.; Qi, Y.; Wilkerson, M.D.; Miller, C.R.; Ding, L.; Golub, T.; Mesirov, J.P.; et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* **2010**, *17*, 98–110.
15. Himar, F.; Samuel, O.; Silvester, K.; Callico, G.M.; Diederik, B.; Adam, S.; Pineiro, J.F.; Roberto, S. HELICoiD project: A new use of hyperspectral imaging for brain cancer detection in real-time during neurosurgical operations. In Proceedings of the SPIE Commercial+ Scientific Sensing and Imaging: International Society for Optics and Photonics, Baltimore, MD, USA, 10 May 2016.
16. Pereira, S.; Pinto, A.; Alves, V.; Silva, C.A. Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Trans. Med. Imaging* **2016**, *35*, 1240–1251.
17. Tate, A.R.; Majós, C.; Moreno, A.; Howe, F.A.; Griffiths, J.R.; Arús, C. Automated classification of short echo time in in vivo 1H brain tumor spectra: a multicenter study. *Magn. Reson. Med.* **2003**, *49*, 29–36.
18. Majós, C.; Julià-Sapé, M.; Alonso, J.; Serrallonga, M.; Aguilera, C.; Acebes, J.J.; Gili, J. Brain tumor classification by proton MR spectroscopy: Comparison of diagnostic accuracy at short and long TE. *Am. J. Neuroradiol.* **2004**, *25*, 1696–1704.
19. Ranjith, G.; Parvathy, R.; Vikas, V.; Chandrasekharan, K.; Nair, S. Machine learning methods for the classification of gliomas: Initial results using features extracted from MR spectroscopy. *Neuroradiol. J.* **2015**, *28*, 106–111.
20. Corso, J.J.; Sharon, E.; Dube, S.; El-Saden, S.; Sinha, U.; Yuille, A. Efficient multilevel brain tumor segmentation with integrated bayesian model classification. *IEEE Trans. Med. Imaging* **2008**, *27*, 629–640.
21. Liu, J.; Udupa, J.K.; Odhner, D.; Hackney, D.; Moonis, G. A system for brain tumor volume estimation via MR imaging and fuzzy connectedness. *Comput. Med. Imaging Graph.* **2005**, *29*, 21–34.
22. Kaus, M.R.; Warfield, S.K.; Nabavi, A.; Black, P.M.; Jolesz, F.A.; Kikinis, R. Automated segmentation of MR images of brain tumors. *Radiology* **2001**, *218*, 586–591.
23. Warfield, S.K.; Kaus, M.; Jolesz, F.A.; Kikinis, R. Adaptive template moderated spatially varying statistical classification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Berlin/Heidelberg, Germany, 1998.
24. Vinitzki, S.; Gonzalez, C.F.; Knobler, R.; Andrews, D.; Iwanaga, T.; Curtis, M. Fast tissue segmentation based on a 4D feature map in characterization of intracranial lesions. *J. Magn. Reson. Imaging* **1999**, *9*, 768–776.

25. Prastawa, M.; Bullitt, E.; Moon, N.; Van Leemput, K.; Gerig, G. Automatic brain tumor segmentation by subject specific modification of atlas priors. *Acad. Radiol.* **2003**, *10*, 1341–1348.
26. Ho, S.; Bullitt, E.; Gerig, G. Level-set evolution with region competition: Automatic 3-D segmentation of brain tumors. In Proceedings of the Object Recognition Supported by User Interaction for Service Robots, Quebec, QC, Canada, 11–15 August 2002.
27. Gerig, G.; Jomier, M.; Chakos, M. Valmet: A new validation tool for assessing and improving 3D object segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2001*; Springer: Berlin/Heidelberg, Germany, 2001.
28. Zacharaki, E.I.; Wang, S.; Chawla, S.; Soo Yoo, D.; Wolf, R.; Melhem, E.R.; Davatzikos, C. Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme. *Magn. Reson. Med.* **2009**, *62*, 1609–1618.
29. Lawrence, S.; Giles, C.L.; Tsoi, A.C.; Back, A.D. Face recognition: A convolutional neural-network approach. *IEEE Trans. Neural Netw.* **1997**, *8*, 98–113.
30. Pérez-Carrasco, J.A.; Zhao, B.; Serrano, C.; Acha, B.; Serrano-Gotarredona, T.; Chen, S.; Linares-Barranco, B. Mapping from frame-driven to frame-free event-driven vision systems by low-rate rate coding and coincidence processing—application to feedforward ConvNets. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 2706–2719.
31. Krizhevsky, A.; Ilya, S.; Geoffrey, E.H. Imagenet classification with deep convolutional neural networks. In Proceedings of the Advances in Neural Information Processing Systems (NIPS 2012), Lake Tahoe, NV, USA, 3–6 December 2012.
32. Scarpace, L.; Flanders, A.E.; Jain, R.; Mikkelsen, T.; Andrews, D.W. Data From REMBRANDT. The Cancer Imaging Archive. 2015. Available online: <http://doi.org/10.7937/K9/TCIA.2015.588OZUZB> (accessed on 1 December 2016).
33. Clark, K.; Vendt, B.; Smith, K.; Freymann, J.; Kirby, J.; Koppel, P.; Moore, S.; Phillips, S.; Maffitt, D.; Pringle, M.; et al. The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. *J. Digit. Imaging* **2013**, *26*, 1045–1057.
34. Li, Z.; Yu, T.; Hu, G.Z.; Yu, X. Multiple gliomas. *Chin. J. Clin. Oncol.* **2007**, *4*, 379–383.
35. Coates, A.; Andrew, N.; Honglak, L. An analysis of single-layer networks in unsupervised feature learning. In Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, Fort Lauderdale, FL, USA, 11–13 April 2011.
36. Powers, D.M. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *J. Mach. Learn. Technol.* **2011**, *2*, 37–63.
37. Moayedikia, A.; Ong, K.L.; Boo, Y.L.; Yeoh, W.G.; Jensen, R. Feature selection for high dimensional imbalanced class data using harmony search. *Eng. Appl. Artif. Intell.* **2017**, *57*, 38–49.
38. Soltaninejad, M.; Yang, G.; Lambrou, T.; Allinson, N.; Jones, T.L.; Barrick, T.R.; Franklyn, A.H.; Ye, X. Automated brain tumour detection and segmentation using superpixel-based extremely randomized trees in FLAIR MRI. *Int. J. Comput. Assist. Radiol. Surg.* **2017**, *12*, 183–203.
39. Shah, G.D.; Kesari, S.; Xu, R.; Batchelor, T.T.; O’Neill, A.M.; Hochberg, F.H.; Levy, B.; Bradshaw, J.; Wen, P.Y. Comparison of linear and volumetric criteria in assessing tumor response in adult high-grade gliomas. *Neurooncology* **2006**, *8*, 38–46.

