



A Novel Deep Learning Algorithm for the Automatic Detection of High-Grade Gliomas on T2-Weighted Magnetic Resonance Images: A Preliminary Machine Learning Study

Mehmet Ali ATICI¹, Seref SAGIROGLU¹, Pinar CELTIKCI², Murat UCAR³, Alp Ozgun BORCEK⁴, Hakan EMMEZ⁵, Emrah CELTIKCI⁵

¹Gazi University Faculty of Engineering, Department of Computer Engineering, Ankara, Turkey

²Baskent University Faculty of Medicine, Department of Radiology, Ankara, Turkey

³Gazi University Faculty of Medicine, Department of Radiology, Ankara, Turkey

⁴Gazi University Faculty of Medicine, Department of Neurosurgery, Division of Pediatric Neurosurgery, Ankara, Turkey

⁵Gazi University Faculty of Medicine, Department of Neurosurgery, Ankara, Turkey

Corresponding author: Emrah CELTIKCI ✉ drceltikci@gmail.com

ABSTRACT

AIM: To propose a convolutional neural network (CNN) for the automatic detection of high-grade gliomas (HGGs) on T2-weighted magnetic resonance imaging (MRI) scans.

MATERIAL and METHODS: A total of 3580 images obtained from 179 individuals were used for training and validation. After random rotation and vertical flip, training data was augmented by factor of 10 in each iteration. In order to increase data processing time, every single image converted into a Jpeg image which has a resolution of 320x320. Accuracy, precision and recall rates were calculated after training of the algorithm.

RESULTS: Following training, CNN achieved acceptable performance ratios of 0.854 to 0.944 for accuracy, 0.812 to 0.980 for precision and 0.738 to 0.907 for recall. Also, CNN was able to detect HGG cases even though there is no apparent mass lesion in the given image.

CONCLUSION: Our preliminary findings demonstrate; currently proposed CNN model achieves acceptable performance results for the automatic detection of HGGs on T2-weighted images.

KEYWORDS: Artificial intelligence, Deep learning, Glioma, Machine learning, Magnetic resonance imaging

INTRODUCTION

High grade gliomas (HGGs) are tumors of the central nervous system (CNS) with poor prognosis and low survival rates (12). Age-adjusted overall incidence of HGGs is 5.91 per 100,000 person, in the United States of America (3). Magnetic resonance imaging (MRI) is the first-line imaging modality in the diagnosis and follow-up of HGGs (15).

According to the data from Turkish Society of Radiology, Turkey has the highest number of MRI studies per person amongst

Organization for Economic Co-operation and Development (OECD) countries (144 per 1000 person). On the other hand, the number of radiologists per 100,000 person is only five and average time for evaluating and reporting radiological studies is approximately five minutes (<https://www.turkrad.org.tr/>). Without proper triage and arrangement of brain MRI studies prior to evaluation and reporting, this huge workload may result in delayed, misdiagnosis or even missed diagnosis of HGGs as well as other CNS pathologies.

Mehmet Ali ATICI : 0000-0002-0673-5724
Seref SAGIROGLU : 0000-0003-0805-5818
Pinar CELTIKCI : 0000-0002-1655-6957

Murat UCAR : 0000-0003-2650-0065
Alp Ozgun BORCEK : 0000-0002-6222-382X
Hakan EMMEZ : 0000-0002-3290-179X

Emrah CELTIKCI : 0000-0001-5733-7542

In deep learning, a convolutional neural network (CNN) is a class of deep neural networks, most commonly applied to analyze visual imagery. Automated pre-evaluation of the MRI studies by a CNN prior to a human specialist, could provide faster diagnosis of HGGs via generating a warning to the clinician or providing priority to MRI studies with tumors in the Picture Archiving and Communication System (PACS).

In this study, we aimed to solve this problem by proposing and training a novel deep learning algorithm, a CNN for the automatic detection of HGGs on T2-weighted MRI images. We believe this would provide faster and complete diagnosis of these tumors, besides easing the workload of radiologists.

MATERIAL and METHODS

Patients

This study was performed as a part of joint project between Department of Neurosurgery of Faculty of Medicine and Department of Computer Engineering of Faculty of Engineering at Gazi University, using the laboratory facilities of Gazi University Big Data and Information Security Center Research

Lab (BIDISEC) and Brain Research Study Groups. Dataset used in this study was comprised of normal subjects from the Radiology Department of Gazi University Faculty of Medicine and HGG patients from the Neurosurgery Department of Gazi University Faculty of Medicine. Prior to study, local ethics committee approval was obtained as well as written informed consent of all subjects.

Data Processing

In order to achieve the task automatically, a model based on artificial neural networks is established according to the data sets. In order to establish a ground truth, a team of radiologists and neurosurgeons evaluated and classified every single image in the dataset as normal subject and HGG case. All imaging data was then preprocessed and prepared for the CNN model. CNN was initially trained with each image in T2-weighted series to detect whether the image belongs to a normal subject or an HGG case. All 20 images of the T2-weighted series were numbered from S0 to S19 for both groups (Figures 1, 2). Workflow of the data processing is summarized in Figure 3. In summary; 1) Digital Imaging and Communications in Medicine (DICOM) file was loaded from

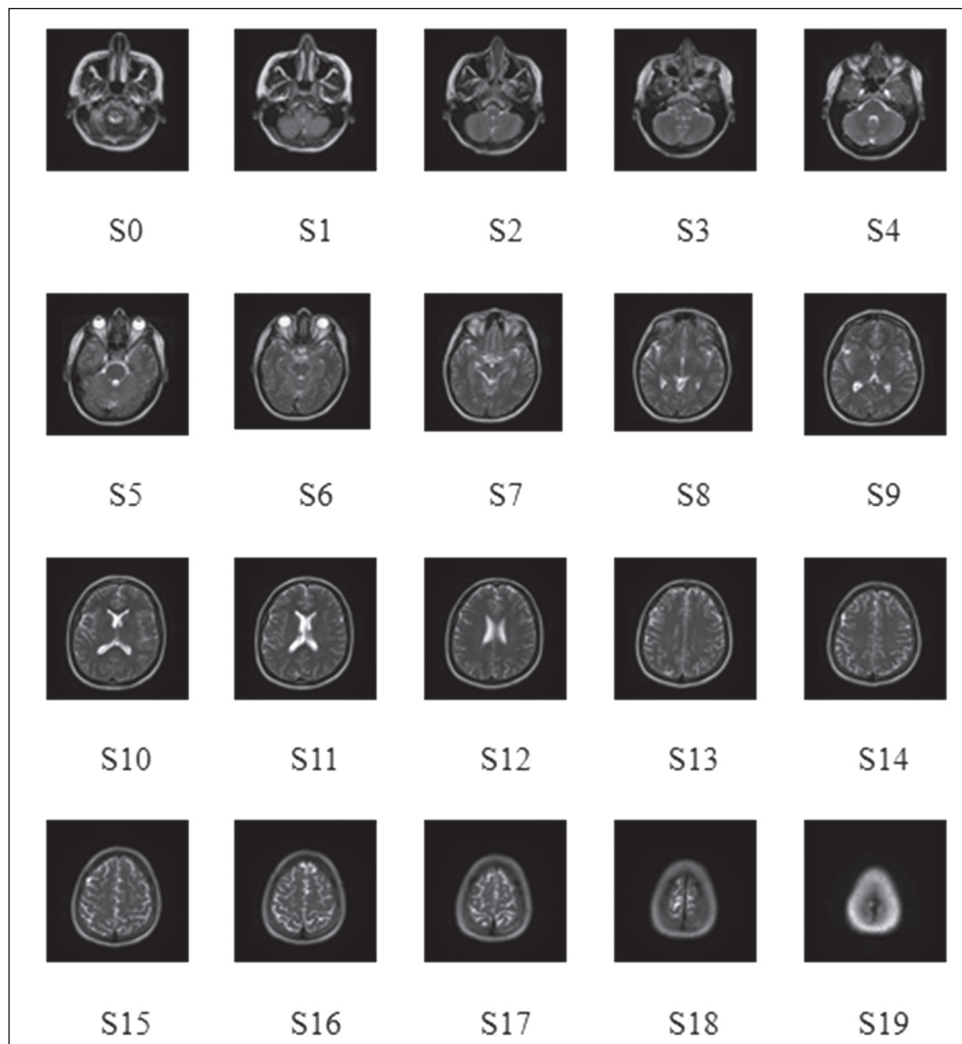


Figure 1: Sample images of T2-weighted images sequence for a random normal subject.

the repository. 2) DICOM file was parsed, pixel data, bit depth, max pixel value and min pixel value was extracted. 3) Pixel values were converted to 8 bits by using the values extracted in the previous step. 4) 8 bits pixel array was converted to Jpeg image. 5) Jpeg image was resized to a lower resolution of 320x320 dpi in order to provide faster data processing. If width and height of image was not equal to this resolution, image was filled with 0 values symmetrically to fit the size exactly to resolution of 320x320 dpi. 6) Finally, Jpeg image was converted to gray scale and given as input to CNN.

CNN Architecture

Our custom CNN architecture, which is a slightly different version of the model proposed in the source (https://github.com/CShorten/KaggleDogBreedChallenge/blob/master/DogBreed_BinaryClassification.ipynb). The number of convolutions of these blocks are 32, 64 or 96. We use 3x3 filters in convolution layers and 2x2 filters in pooling layers. Batch normalization is the process of normalizing inputs of intermediate layers to avoid internal covariate shift problem

(10). In batch normalization, following three steps were conducted in order: 1) Mean and variance of the mini batch was calculated, 2) input of the layers was normalized by using mean and variance, and 3) scale and shift applied to produce the output of the layer.

The architecture of the entire network is given in Figures 4 and 5. Architecture of CNN consists of five convolutional blocks and a fully connected layer. In fully connected layer, the output of the last convolutional block was flattened and fully connected to first dense layer. We have also used dropout layer between dense layers to regularize the network for reducing over fitting. In dropout technique, some random nodes were shut down and not utilized in back-propagation and feed forward calculations. The analyses was conducted on basis of 5-fold cross validation on the prepared dataset by using Keras (<https://github.com/fchollet/keras>); a deep learning framework in Python (Python Software Foundation, <https://www.python.org/>). In each iteration, training data was augmented by random rotation in both clockwise and

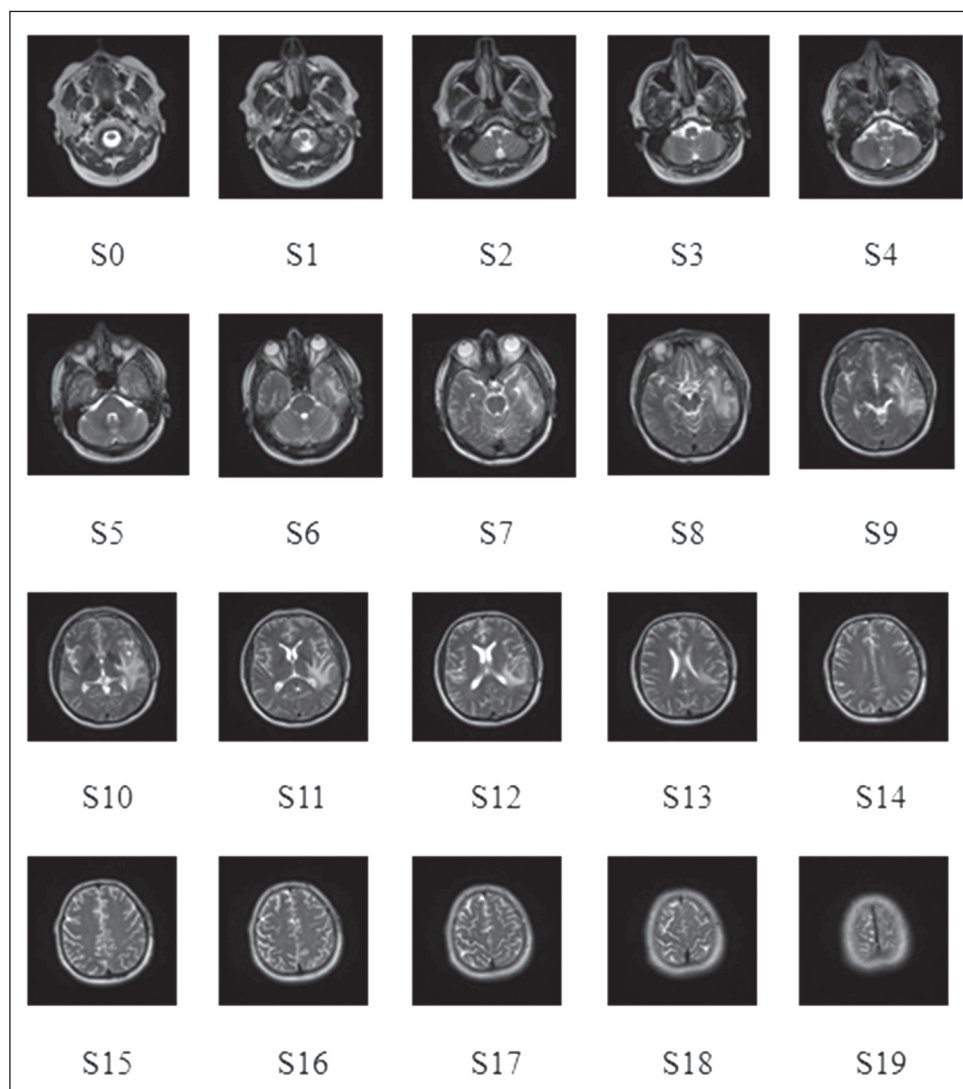


Figure 2: Sample images of T2-weighted images sequence for high grade glioma case.

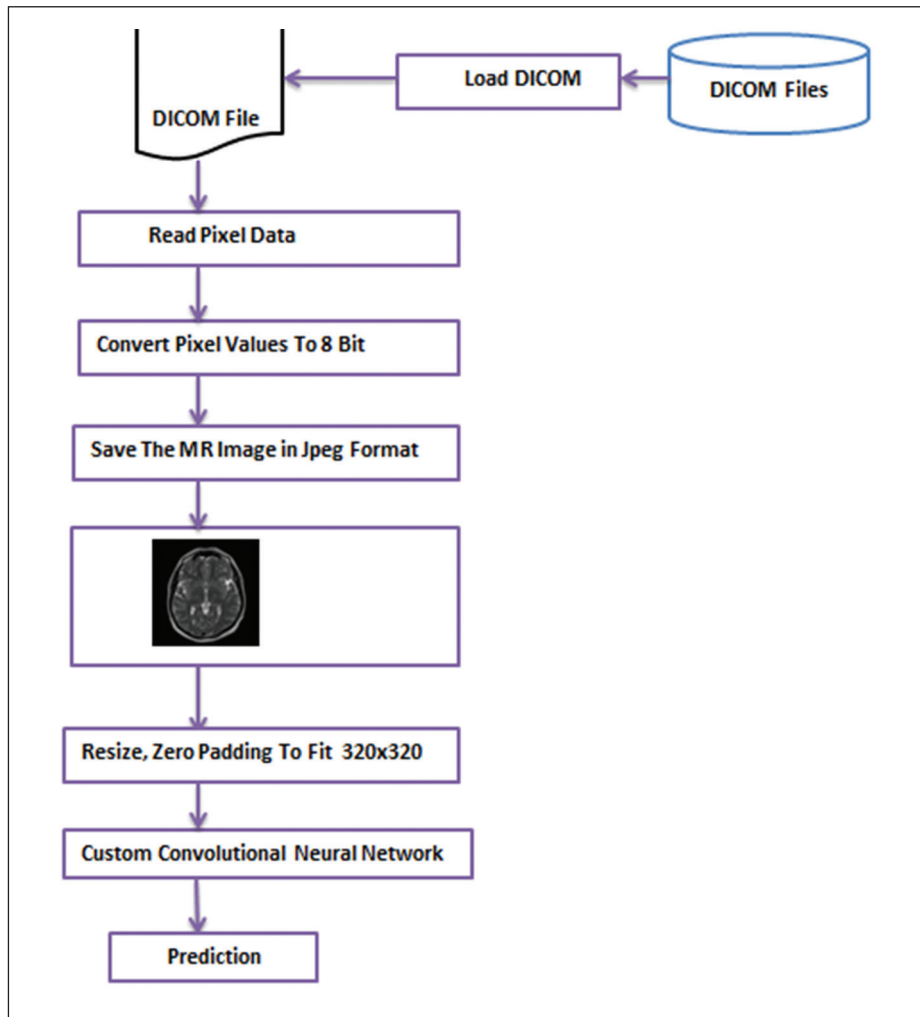


Figure 3: Workflow of proposed convolutional neural network model.

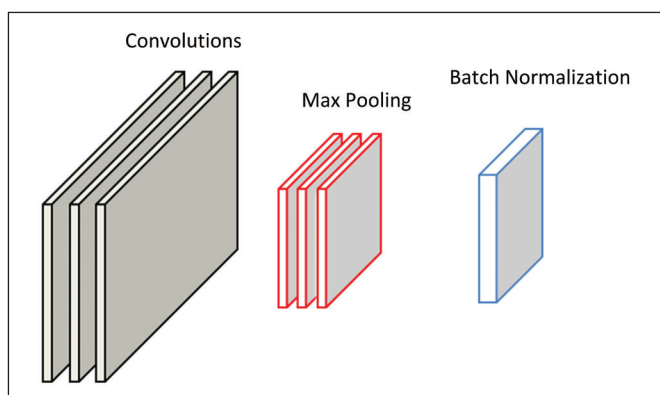


Figure 4: Structure of the Convolution Block in the proposed model.

counterclockwise directions. As T2-weighted images of brain are almost symmetrical, training data was also augmented by vertical flip.

Validation

In order to evaluate the performance of the proposed model,

precision and recall measures were calculated with respect to confusion matrix (5) and given in Table I. Precision, recall and F-measure values can be formally defined as; precision = TP / (TP+FP), recall = sensitivity = true positive rate = TP / (TP+FN), accuracy = (TP+TN) / (TP+FN+TN+FP) (5). Precision means the percentage of your results which are relevant. On the other hand, recall refers to the percentage of total relevant results correctly classified by your algorithm.

RESULTS

T2-weighted MRI images of 114 normal subjects and 65 histopathologically proven HGG cases were used as the dataset. For each network, there was a total of 3580 images for training and validation. Following random rotation and vertical flip, training data was augmented by factor of 10 in each iteration.

Following training, CNN achieved acceptable performance ratios of 0.854 to 0.944 for accuracy, 0.812 to 0.980 for precision and 0.738 to 0.907 for recall (Table II). K-fold cross validation provides more realistic results according to random splitting. For example, if we coincide a random data split for

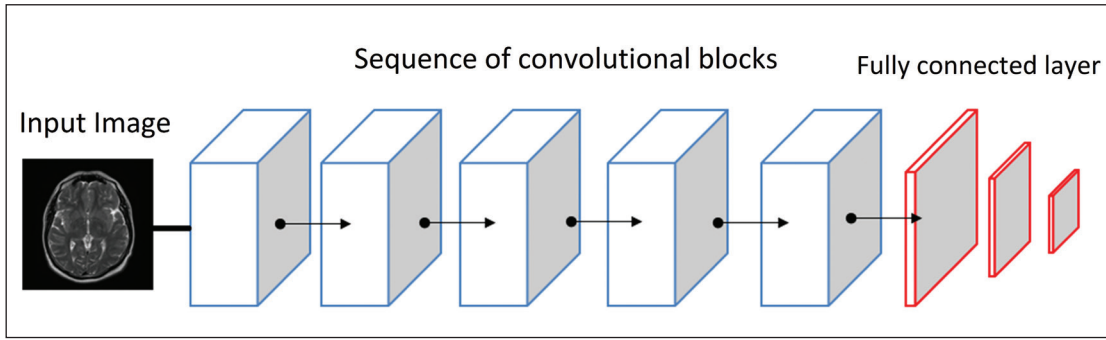


Figure 5: Architecture of the convolutional neural network.

Table I: Confusion Matrix

	Predicted Normal	Predicted GBM
Actual Normal	TN	FP
Actual GBM	FN	TP

TN: True negative, FP: False positive, FN: False negative , TP: True positive.

Table II: Experimental Results

	TN	FN	TP	FP	Accuracy	Precision	Recall
S0	110	10	55	4	0.921	0.932	0.846
S1	109	10	55	5	0.916	0.916	0.846
S2	105	17	48	9	0.854	0.842	0.738
S3	113	15	50	1	0.910	0.980	0.769
S4	108	13	54	6	0.905	0.900	0.830
S5	111	15	50	4	0.899	0.925	0.769
S6	111	9	56	3	0.932	0.949	0.861
S7	111	7	58	3	0.944	0.950	0.892
S8	110	10	55	4	0.921	0.932	0.846
S9	112	9	56	2	0.938	0.965	0.861
S10	110	9	56	4	0.927	0.933	0.861
S11	110	10	55	4	0.921	0.932	0.846
S12	106	15	50	8	0.871	0.862	0.769
S13	104	7	58	10	0.905	0.852	0.892
S14	113	16	49	1	0.905	0.980	0.753
S15	108	9	56	6	0.916	0.903	0.861
S16	105	10	55	9	0.893	0.859	0.846
S17	107	10	55	7	0.905	0.887	0.846
S18	102	13	52	12	0.916	0.812	0.800
S19	106	6	59	8	0.921	0.880	0.907

TN: True negative, FP: False positive, FN: False negative , TP: True positive.

training and testing like fold-5 in Table III, high performance ratios could be achieved. However, these results would be misleading according to the average performance of k-fold cross validation. One of the unexpected achievements of CNN in this study was, CNN was able to detect HGG cases even though there is no mass lesion in the given image. We believe the reason behind is that CNN was able to detect the accompanying changes such as the edema or shifting due to HGG, as well as the tumor itself.

■ DISCUSSION

Here, we propose a novel deep learning approach for the automated detection of HGGs in T2-weighted brain MRI images, with acceptable performance rates. In addition, our model was able to detect anomalies even when given a single image without any apparent lesion. The experimental results indicate that CNN can learn the discriminative patterns of HGG cases which may be overlooked by human experts. This feature could facilitate the successful detection of HGG cases by using any of T2-weighted images in the sequence.

CNNs were created by mimicking the network pattern between neurons of the visual cortex (6–8). A single neuron responds to the stimuli of a limited area in the receptive field. Receptive fields of individual neurons cover the entire visual field via overlapping. CNNs operate similar to this principle. In comparison to other image classification algorithms, CNNs require less pre-processing due to fact that a CNN is able to learn the filters that a conventional algorithm would need human intervention. This provides a major advantage and independence because there is no need of previous knowledge or a human operator.

Anaraki et al. proposed a CNN to classify different grades of gliomas. In this work, they optimized the network architecture by generic algorithms (1). They used publicly available datasets for training and augmented the collected data by rotation and scaling. In experiments, the proposed network achieved 90.9% classification accuracy for three glioma grades and 94.2% accuracy for classification of glioma, meningioma, and pituitary tumors. Also, Hussain et al. designed a CNN for automated brain tumor segmentation which utilized MRI sequences and proposed a two-phase weighted method for training (9). They used batch normalization and dropout layers

for regularizing the network architecture. BRATS 2013 and BRATS 2015 datasets were used in their study and achieved improved results compared to similar works. Also, there are other studies in literature applying CNN algorithms that achieved accuracy rates between 88% and 98.6% (13,14,16). Our results have acceptable accuracy rates similar to these studies. On the other hand, we did not use publicly available standardized datasets. Instead we utilized a dataset which consists of normal subjects and patients with histopathological diagnosis of HGG.

Automatic detection of brain abnormalities would present with many advantages and convenience in the clinical practice such as; informing the clinician earlier than the radiological report, and reducing the rates of missed and misdiagnosis. It is already known that computer-aided diagnosis algorithms assist in reducing diagnostic errors (4,11). Additionally, human experts have an intra-observer variability which means they could have different decisions in different times. Whereas, an algorithm is deterministic and intra-observer variability is non-existent (2).

Our study has several limitations. First, these are the preliminary results of a project which is still in data collection process. Therefore, sample size is limited and the algorithm requires augmentation of training data. Moreover, prior to testing, all images were converted to a lower resolution jpeg format in order to decrease processing time. This procedure decreases the amount of data supplied to the algorithm which may have a negative effect on the performance. Additionally, our study included T2-weighted brain MRI images of both normal subjects and HGG patients. Not including other sequences such as contrast enhanced T1-weighted brain MRI images would decrease precision, sensitivity and accuracy rates. However, our results demonstrated that even only with T2-weighted brain MRI images, our algorithm was successful enough to achieve acceptable results.

■ CONCLUSION

Our preliminary findings demonstrate; currently proposed CNN model is capable of achieving acceptable performance results for the automatic detection of HGGs on T2-weighted images. Additionally, we have shown that CNN could distinguish HGG cases from normal subjects, even from a single MRI image

Table III: Experimental Results for S1 MRI

	TN	FN	TP	FP	Accuracy	Precision	Recall
Fold-1	22	3	9	2	0.861	0.818	0.750
Fold-2	22	4	10	0	0.888	1	0.714
Fold-3	22	1	13	0	0.972	1	0.928
Fold-4	19	2	13	2	0.888	0.866	0.866
Fold-5	24	0	10	1	0.971	0.909	1
Total	109	10	55	5	0.916	0.916	0.846

TN: True negative, **FP:** False positive, **FN:** False negative, **TP:** True positive.

without any apparent lesion. Encouraged with the promising results of this preliminary study; next we aim to provide automatic detection without converting the files to a lower resolution, achieve acceptable performance rates without providing ground truths, automatic segmentation of gliomas, automatic characterization of glioma subtypes and automatic prediction of survival from MRI images in future studies.

■ REFERENCES

- Anaraki AK, Ayati M, Kazemi F: Magnetic resonance imaging-based brain tumor grades classification and grading via convolutional neural networks and genetic algorithms. *Biocybern. Biomed Eng* 39:63–74, 2019
- Buda M, Saha A, Mazurowski MA: Association of genomic subtypes of lower-grade gliomas with shape features automatically extracted by a deep learning algorithm. *Comput Biol Med* 109:218–225, 2019
- Chien LN, Gittleman H, Ostrom QT, Hung KS, Sloan AE, Hsieh YC, Kruchko C, Rogers LR, Wang YFG, Chiou HY, Barnholtz-Sloan JS: Comparative brain and central nervous system tumor incidence and survival between the united states and taiwan based on population-based registry. *Front Public Health* 4:151, 2016
- El-Kareh R, Hasan O, Schiff GD: Use of health information technology to reduce diagnostic errors. *BMJ Qual Saf* 22: ii40–ii51, 2013
- Fawcett T: An introduction to ROC analysis. *Pattern Recognit Lett* 27:861–874, 2006
- Fukushima K: Neocognitron. *Scholarpedia* 2:1717, 2007
- Fukushima K: Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol Cybern* 36:193–202, 1980
- Hubel DH, Wiesel TN: Receptive fields and functional architecture of monkey striate cortex. *J Physiol* 195:215–243, 1968
- Hussain S, Anwar SM, Majid M: Segmentation of glioma tumors in brain using deep convolutional neural network. *Neurocomputing* 282:248–261, 2018
- Ioffe S, Szegedy C: Batch normalization: Accelerating deep network training by reducing internal covariate shift [Internet]; in *International Conference on Machine Learning*. 2015 [cited 2019 May 12]. Available from: <http://proceedings.mlr.press/v37/loff15.html>
- Kooi T, Litjens G, van Ginneken B, Gubern-Mérida A, Sánchez CI, Mann R, den Heeten A, Karssemeijer N: Large scale deep learning for computer aided detection of mammographic lesions. *Med Image Anal* 35:303–312, 2017
- Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, Ohgaki H, Wiestler OD, Kleihues P, Ellison DW: The 2016 World Health Organization Classification of Tumors of the Central Nervous System: A summary. *Acta Neuropathol (Berl)* 131:803–820, 2016
- Lu S, Lei L, Huang H, Xiao L: A hybrid extraction-classification method for brain segmentation in MR image; in *2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. 2016
- Narayana PA, Coronado I, Robinson M, Sujit SJ, Datta S, Sun X, Lublin FD, Wolinsky JS, Gabr RE: Multimodal MRI Segmentation of Brain Tissue and T2-Hyperintense White Matter Lesions in Multiple Sclerosis using Deep Convolutional Neural Networks and a Large Multi-center Image Database; in *2018 9th Cairo International Biomedical Engineering Conference (CIBEC)*. 2018
- Tabatabai G, Stupp R, van den Bent MJ, Hegi ME, Tonn JC, Wick W, Weller M: Molecular diagnostics of gliomas: The clinical perspective. *Acta Neuropathol (Berl)* 120:585–592, 2010
- Zhang B, Chang K, Ramkissoon S, Tanguturi S, Bi WL, Reardon DA, Ligon KL, Alexander BM, Wen PY, Huang RY: Multimodal MRI features predict isocitrate dehydrogenase genotype in high-grade gliomas. *Neuro-Oncol* 19:109–117, 2017