DOI Number: 10.5958/0974-1283.2019.00132.4

Detection of HGG and LGG Brain Tumors using U-Net

Sang-Geun Choi[1], Chae-Bong Sohn[2]

[1]Ph.D. Candidate, [2]Professor, Department of Electronics and Communications Engineering, Kwangwoon University, 20 Kwangwoon-Ro, Nowon-Gu, Seoul, 01897, Korea

ABSTRACT

Background/Objectives: Advancement in medical equipment has enabled accurate and quick diagnosis in medical field. However, an increase in the number of medical staff is slower than the rate of medical equipment development. It has resulted in increased risk of diagnostic misinterpretation. The purpose of this paper is to help diagnosis of medical staff through artificial neural network (ANN).

Method/Statistical Analysis: We selected U-Net among artificial neural networks. U-Net is highly accurate in medical imaging. The dataset for learning the network was obtained from the Brain Tumor Segmentation Challenge (BraTS). This dataset contains four classes of brain tumor data and it is suitable for learning variety of brain tumors. We used F-Score to measure the accuracy of the learned network.

Findings: In this paper, we compare the performance of the network by conducting two experiments. First, we checked the learning progress of the network. Second, we compared the results of learning with mixed and single datasets. In the first experiment, when allowing the network to learn for a total of 200 generations, it was confirmed that the results of 100 generations were the most accurate. In the second experiment, the network learned by three groups of datasets. The first group consisted of HGG data only, and the second group was composed of LGG data only, and the last group was made up of mixing HGG and LGG data. When comparing the results of the first group with the third group, the accuracy of HGG patient was 0.6696 and 0.6222, respectively. Subsequently, the results of the second and the third group were 0.6315 and 0.6228, respectively.

Improvements/Applications: In this experiment, we compared the results obtained when the datasets were mixed and when they were used singly. The results show similar accuracy. However, when using a mixture of datasets, the accuracy is lower, which is enough to assist the diagnosis of the medical staff. It is expected that this will help the development of the medical image processing field by confirming the position and size of the brain tumor accurately regardless of the data of any grade for brain tumor.

Keywords: *Artificial Neural Network (ANN), U-Net, Brain Tumor Segmentation Challenge (BraTS), Brain Tumor, HGG, LGG*

Introduction

Artificial neural networks (ANN) are applied in various fields. The medical field also shows a lot of activities using artificial neural networks. They are trying to prevent the occurrence of patients by collecting biometric information such as patient's life pattern, blood pressure, and body temperature [1]. In the field of medical imaging, there is an active movement for application of artificial neural network to various medical equipments. Ultrasound, computed tomography (CT), magnetic resonance imaging (MRI) and other devices have helped medical staff to make a diagnosis. However, with the development of these devices, the number of patients using them is increasing. In comparison, the number of medical staffs is limited. The possibility of misdiagnosis due to inaccurate reading caused by these issues has emerged as a serious problem. To solve these problems, there have been many attempts by introducing a medical image reading system through artificial neural network[2].

In this paper, we apply artificial neural network to brain tumor detection to help more accurate diagnosis in

Corresponding Author:
Chae-Bong Sohn
Professor, Department of Electronics and Communications Engineering,
Kwangwoon University, Korea
Email: cbsohn@kw.ac.kr

MRI image. We selected U-Net [3], which is widely used in medical imaging field. The learning of the network was set to use the dataset provided by the Brain Tumor Detection Challenge (BraTS) [4]. This dataset provides two main classes of brain tumors. Learning will proceed with the configuration of the dataset in two ways. The first one learns only a single grade of brain tumor and the second one learns regardless of the brain tumor grade.

In Chapter 2, we describe the network and dataset used in learning, the overall method of experiment, and the evaluation method to analyze the results. In Chapter 3, we will identify the performance and results of the network and analyze the results. Finally, in Chapter 4, we will present the expected effects and future directions when using this paper.

Materials and Method

U-Net: Common CNN (Convolutional Neural Network) is used for classification. In this case, the label for the class is output as a result of the input image. However, localization results are required in image processing such as medical imaging. The reason we selected U-Net is as follow. First, this network is localizable and can detect tumor location and size. Second, to solve the problem caused by a small amount of training data, the learning is performed by dividing the image into patches [3].

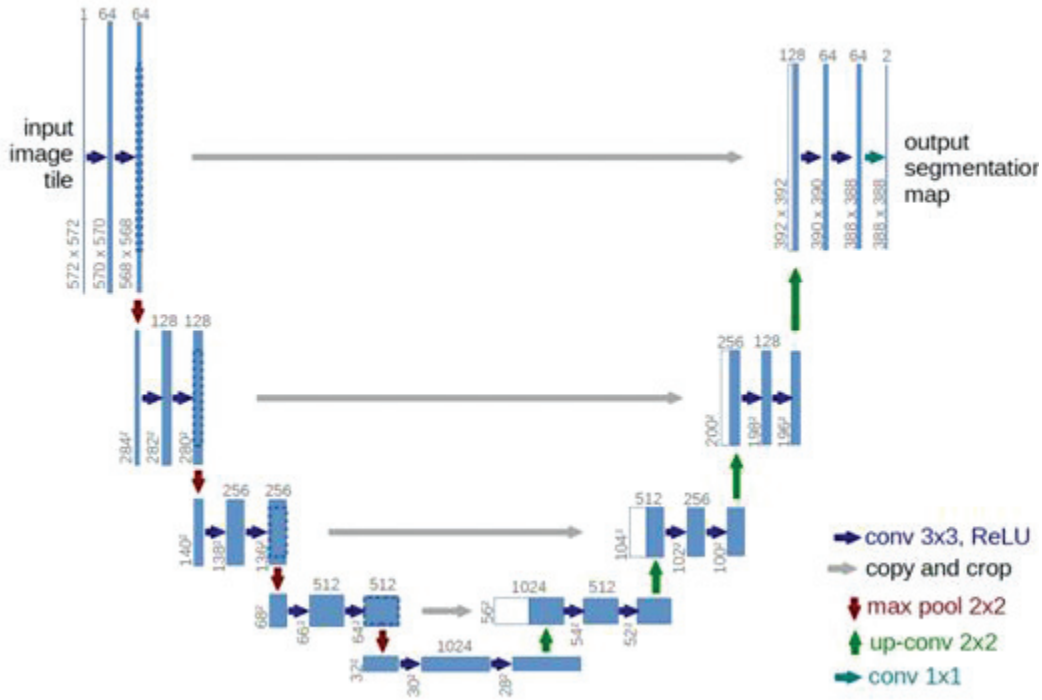

Figure 1: U-Net architecture

Network Architecture: Figure 1 shows the overall architecture of U-Net. It consists of a contracting path in left side and an expansive path in right side of Figure 1. The contracting path corresponding to down-sampling proceeds by repeating two convolution layer and max pooling operation. The feature channel doubles when down-sampling is performed. The expansive path corresponding to up-sampling is like the contracting path, but there are two differences. First, up-convolution is performed instead of max pooling. This reduces the feature channel in half and doubles the feature map. Second, when up-sampling is performed, the feature map is copied and cropped in the contracting path to connect. Crop is required because the size of the corresponding feature map is smaller in the expansive path. This structure allows more accurate localization. This part was added in CNN, allowing U-Net to enable semantic segmentation. Finally, one by one convolution is performed to map the feature vector to the desired class.

Datasets for Training: The MRI image datasets used for training was a dataset provided by Brain Tumor Segmentation Challenge (BraTS) [4]. These datasets consist of 220 sets of HGG data and 54 sets of LGG data. Each MRI set consists of four sets of images with different contrast: T1, T2, FLAIR and T1c. This image set consists of 150 brain tomographic images.

HGG and LGG: Gliomas are characterized by subtypes and by a numerical grading system. HGG and LGG are subtypes of glioma. The grade of a tumor means how the cancer cells appear under a microscope. Grade I tumors grow slowly and can sometimes be totally removed by surgery, while grade IV tumors are fast-growing, aggressive and difficult to treat [5,6,7].

According to the current World Health Organization (WHO) scheme, malignant astrocytomas are classified and graded as shown in Table 1.

Table 1: Grade chart of glioma

	Grade	Comments
LGG	I	Benign, slow-growing tumor Usually associated with long term survival Least likely to recur
	II	Increased hypercellularity No mitosis No vascular proliferation No necrosis Can recur as a higher-grade tumor
HGG	III	High rate of hypercellularity High rate of mitosis No vascular proliferation No necrosis High rate of tumor recurrence
	IV	Very high rate of hypercellularity Very high rate of mitosis Presence of vascular proliferation Presence of necrosis

For four grades of glioma, it can classify grade I and II as LGG and grade III and IV as HGG. Figure 2 shows MRI images of the brain corresponding to HGG and LGG, which are included in the dataset used for training.

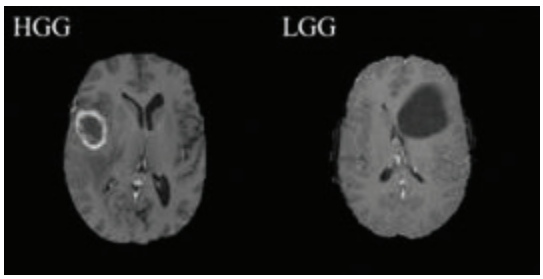

Figure 2: Brain MRI of patients with HGG and LGG

T1, T2, FLAIR and T1c: MRI images can acquire images of different contrasts according to different techniques or acquisition parameters. It is possible to emphasize the desired area by adjusting the value of the weight or adjusting the shooting direction [8,9]. The dataset provided by BraTS consists of the following four images as shown in Table 2.

Table 2: Type of MRI images

Type	Feature
T1	T1-weighted MRI Measure T1 (longitudinal) relaxation time of tissue Tissue with short relaxation times are brighter
T2	T2-weighted MRI Measure T2 (transverse) relaxation time of tissue Tissue with long relaxation times are brighter
T1c	T1-weighted MRI after administration of contrast agent The signal for tumor increase
FLAIR	Fluid Attenuated Inversion Recovery MRI Bright signal of CSF (Cerebrospinal Fluid) is suppressed Can detect small hyperintense lesions better

The purpose of this paper is to detect tumors. For this reason, the training was performed using T1c-weighted MRI, which administration a contrast agent that enhances the signal of the tumor. Figure 3 shows MRI images of each type [11,12,13].

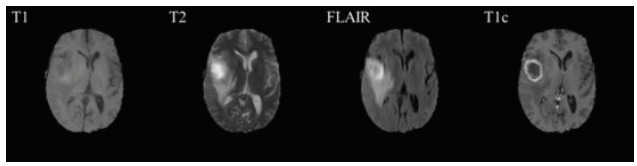

Figure 3: Brain MRI of patients with T1, T2, FLAIR and T1c

Experimental Methods: In this paper, the network will learn dataset of all three configurations. The first network will only have the HGG data, the second will only the LGG data, and finally the HGG and LGG will be learned and the results will be compared. Table 3 shows the number of datasets to be used for learning

Table 3: Amount of Training and Test

	a. HGG only	b. LGG only	c. HGG and LGG
Training	200 sets	50 sets	250 sets
Test	20 sets	4 sets	24 sets

At the same time, experiments were conducted to confirm learning progress of the network. This experiment trained the network over a total of 200 generations and confirmed the change in accuracy. The generations to be used for comparison are 10, 50, 100 and 200 generations of learning data.

F-Score: In this paper, F-Score was chosen to measure the accuracy of the results. For other methods, such as PSNR, which compares image differences, the entire region is compared, and the result is measured. However, in this paper, which is aimed at the detection of brain tumor, the size of non-tumor parts is relatively large when the image of the result is confirmed. Therefore, even if the result does not come out well, there is a problem that the score becomes large. The F-Score can be used to calculate more precisely because it only uses information about the part to be found and the parts that are not correctly found. F-Score can be measured by Equation (1) [10].

$$F = \left(\frac{recall^{-1} + precesion^{-1}}{2} \right)^{-1} = 2 \times \frac{precesion \times recall}{precesion + recall} \quad \ldots(1)$$

In Equation (1), precision and recall are defined by Equation (2).

$$precision = \frac{TP}{TP + FP}, \; recall = \frac{TP}{TP + FN} \quad \ldots(2)$$

Each element of Equation (2) follows the definition of the confusion matrix in Table 4.

Table 4: Confusion Matrix

		True condition	
		Condition Positive	**Condition Negative**
Predicted Condition	Predicted Condition Positive	True Positive (TP)	False Positive (FP)
	Predicted Condition Negative	False Negative (FN)	True Negative (TN)

Results and Discussion

First, we compared the F-Score of each generation to confirm the progress of the network learning. Figure 4 is a chart showing F-Score for each generation of MRI test images. The MRI images used for the test are for HGG patients.

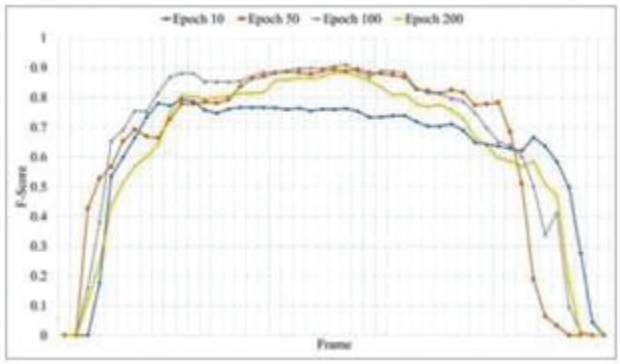

Figure 4: F-score for each epoch

Table 5 shows the average F-Score for each generation.

Table 5: Average of F-score for each epoch

Epoch	10	50	100	200
Average	0.613759	0.637009	0.669652	0.62122
Recall	0.824101	0.701977	0.720442	0.765566
Precision	0.488958	0.583047	0.625551	0.522671

From Figure 4 and Table 5, we can see that the overall network accuracy improves. However, the results of the 200th generation are lower than those of the 50th generation.

The next experiment compares the accuracy of the network with the different configuration of the dataset. The learning was proceeded with the configuration of Table 3. Based on above result, the network learning was carried out over 100 generation. Results were compared between group (a) and group (c) in Table 3, and between group (b) and group (c). MRI images of the patients to be used for test are those corresponding to HGG and LGG brain tumors, respectively. Table 6 shows the result of the comparison.

Table 6: Comparison of three training sets

	HGG Patient			**LGG Patient**		
Trained sets	**F-Score**	**Recall**	**Precision**	**F-Score**	**Recall**	**Precision**
HGG only	0.669652	0.720442	0.625551	-	-	-
LGG only	-	-	-	0.631513	0.757418	0.5415
HGG and LGG	0.622251	0.701455	0.559118	0.622863	0.638845	0.607661

From the results in Table 6, the results of group (c) shows lower accuracy than group (a) and (b). The difference is 0.047401 for HGG patient and 0.00865 for LGG patient.

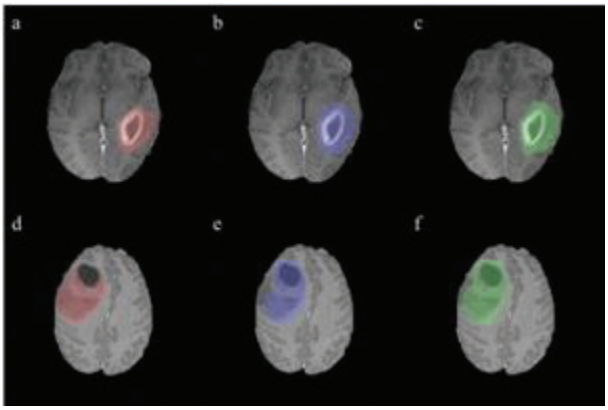

Figure 5: Brain Tumor Detection Result

Figure 5 shows the MRI images of two patients used as testsets in the experiment. The above three images are from HGG patient and the below three images from LGG patient. From left to right, ground truth, the result of learning a single grade of dataset, and the result of learning a mixture of datasets.

Conclusion

In this paper, we compare the learning progress of the network and the accuracy according to the dataset configuration. In first experiment, when the learning was conducted over a total 200 generations, it was confirmed that the accuracy was lowered after 100 generation. This may be due to overfitting or local minima. Indeed, the average of the results of the training set has an F-Score value of 0.7 to 0.8, which is significantly higher than the test results. It is expected that better results will be obtained by adding Dropout to prevent network overfitting.

The following experiment is a comparison of the accuracy according to the configuration of the dataset. The results show that group (c) is 0.02 to 0.04 less accurate than group (a) and (b). This difference can be confirmed in Figure 5. As shown in Figure 5, both networks accurately detect the location of the tumor. However, an error occurred in detecting the size. The difference in results was due to tumor size error. Another problem is that non-tumors are detected as tumors in the results of LGG patients. This problem is due to the similar parts and features of the tumors they exist in the MRI image. Another reason is that overfitting is easily caused by lack of data for LGG patients used for learning.

In conclusion, it is possible to accurately detect the location of the brain tumor and it is expected to be of great help in the diagnosis of the medical staff. Furthermore, it is expected to be useful in various field such as diagnosis of pulmonary tuberculosis by x-ray image and diagnosis of cancer by CT image.

Acknowledgment

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2019-2016-0-00288) supervised by the IITP (Institute for Information & communications Technology Promotion).

Ethical Clearance: Not required

Source of Funding: MSIT

Conflict of Interest: Nil

REFERENCES

1. Baig Mirza Mansoor, Gholamhosseini Hamid. Smart health monitoring systems: an overview of design and modeling. Journal of medical systems, 2013, 37.2: 9898.
2. Shen Dinggang, Wu Guorong, Suk Heung-Il. Deep learning in medical image analysis. Annual review of biomedical engineering, 2017, 19: 221-248.
3. Ronneberger Olaf, Fischer Philipp, Brox Thomas. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015. p. 234-241.
4. Menze Bjoern H, Jakab Andras, Bauer Stefan, Kalpathy-cramer Jayashree, Farahani Keyvan, Kirby Justin, et al. The multimodal brain tumor image segmentation benchmark (BRATS). IEEE transactions on medical imaging, 2015, 34.10: 1993.
5. Theeler Brett J, Groves Morris D. High-grade gliomas. Current treatment options in neurology, 2011, 13.4: 386-399.

6. Stieber Volker W. Low-grade gliomas. Current treatment options in oncology, 2001, 2.6: 495-506.

7. Hakyemez B, Erdogan C, Ercan I, Ergin N, Uysal S, Atahan S. High-grade and low-grade gliomas: differentiation by using perfusion MR imaging. Clinical radiology, 2005, 60.4: 493-502.

8. Hecht M. J, Fellner F, Fellner C, Hilz M. J, Heuss D, Neundorfer B. MRI-FLAIR images of the head show corticospinal tract alterations in ALS patients more frequently than T2-, T1-and proton-density-weighted images. Journal of the neurological sciences, 2001, 186.1-2: 37-44.

9. Merkler Doron, Boretius Susann, Stadelmann Christine, Ernsting Tristan, Michaelis Thomas, Frahm Jens, et al. Multicontrast MRI of remyelination in the central nervous system. NMR in Biomedicine: An International Journal Devoted to the Development and Application of Magnetic Resonance In vivo, 2005, 18.6: 395-403.

10. Sasaki Yutaka. The truth of the F-measure. Teach Tutor mater, 2007, 1.5: 1-5.

11. Jyoti A, Mohanty MN, Mallick P.K. Morphological based segmentation of brain image for tumor detection. InElectronics and Communication Systems (ICECS), 2014 International Conference on 2014 Feb 13 (pp. 1-5). IEEE.

12. Mallick PK,Bhoi AK, Kumar SS, Sherpa KS. Brain Tumor Detection: A Comparative Analysis of Edge Detection Techniques. International Journal of Applied Engineering Research.;10(44):2015.

13. MALLICK PK, MISHRA D, PATANAIK S, SHAW K. A novel supervised gene clustering approachby mining interdependent gene patterns. International Journal of Pharma and Bio Sciences. 2016 Oct;7.