Feasible Study on Intracranial Hemorrhage Detection and Classification using a CNN-LSTM Network

Hoon Ko, Heewon Chung, Hooseok Lee and Jinseok Lee, *Senior Member, IEEE*

Abstract— **Intracranial hemorrhage (ICH) is a life-threatening condition, the outcome of which is associated with stroke, trauma, aneurysm, vascular malformations, high blood pressure, illicit drugs and blood clotting disorders. In this study, we presented the feasibility of the automatic identification and classification of ICH using a head CT image based on deep learning technique. The subtypes of ICH for the classification was intraparenchymal, intraventricular, subarachnoid, subdural and epidural. We first performed windowing to provide three different images: brain window, bone window and subdural window, and trained 4,516,842 head CT images using CNN-LSTM model. We used the Xception model for the deep CNN, and 64 nodes and 32 timesteps for LSTM. For the performance evaluation, we tested 727,392 head CT images, and found the resultant weighted multi-label logarithmic loss was 0.07528. We believe that our proposed method enhances the accuracy of ICH identification and classification and can assist radiologists in the interpretation of head CT images, particularly for brain-related quantitative analysis.**

I. Introduction

An intracranial hemorrhage (ICH) is a type of bleeding inside the brain that accounts for 10 to 20% of all strokes [1-4]. ICH is also considered a critical condition that results in trauma, aneurysm, vascular malformations, high blood pressure, illicit drugs and blood clotting disorders. The incidence of spontaneous ICH is approximately 24.6 per 100,000 worldwide and approximately 40,000 to 67,000 cases per year in the United States [5-8]. The mortality rate ranges from 35% to 52% at 1 month [7]. Thus, the accurate detection of ICH is a critical step in treating the patient.

In clinical practice, trained specialists review medical images of patient's cranium for the investigation of the presence, location and subtype of ICH. However, the process is complicated and often time consuming [9]. Automatic detection of ICH using medical images has the potential to detect ICH earlier and more accurately, leading to improved clinical outcomes. In addition, such a computer-aided diagnosis (CAD) tool helps to improve the performance of radiologists by providing "second-opinion"' in a cost-effect way.

* This study was supported by a Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning NRF-2015M3A9D7067215.

H. Ko, H. Chung, H. Lee and J. Lee are with the Department of Biomedical Engineering, Wonkwang University College of Medicine, Iksan, Jeonbuk, Korea. (corresponding author to provide phone: +82-63-850-6970; e-mail: gonasago@wku.ac.kr).

For the automatic detection of ICH, many computer vision-based studies have been conducted [10-14]. In recent years, deep learning has attracted considerable attention. Deep learning is a data-driven approach for training neural networks to map the input to the desired outputs through nonlinear mathematical functions. Especially, convolutional neural network (CNN) approach has proven effective in a variety of vision tasks such as image classification and object detection [15, 16]. Recently, deep CNN was applied to the automatic detection of ICH with high accuracy identifying even subtle ICH overlooked by radiologists [11].

In this study, we developed a predictive deep learning model using both CNN and long-short term memory (LSTM) networks based on a single head computed tomography (CT) image. In addition, we classified ICH into 5 types such as intraparenchymal, intraventricular, subarachnoid, subdural and epidural. For the identification and classification of ICH, we used Xception model [17] as a backbone, and connected the model to LSTM network followed by dense layer for the final output.

II. Methods

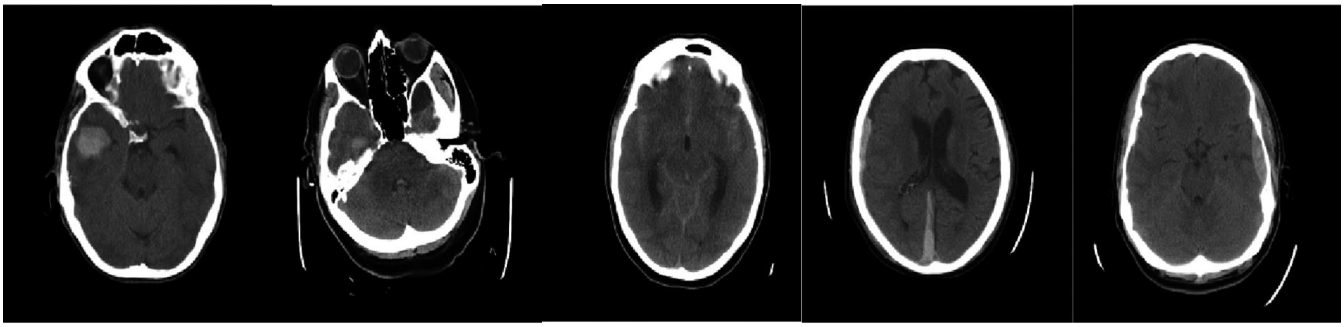
A. Database

The American Society of Neuroradiology (ASNR) and the Radiological Society of North America (RSNA) have teamed up for deep learning challenge to identify ICH and to classify its subtype from brain CT images in 2019. Four research institutions from North and South America provided large volumes of de-identified CT studies that were assembled to create the challenge dataset: Stanford University, Thomas Jefferson University, Unity Health Toronto and Universidade Federal de São Paulo (UNIFESP). During the challenge, researchers and scientists worldwide competed to develop deep neural network to detect ICH and classify its subtype such as intraparenchymal, intraventricular, subarachnoid, subdural and epidural. In the competition rule, anyone may access and use the competition data for non-commercial purposes only, and for academic research and education.

The challenge database includes two datasets: stage-1 and stage-2. After the challenge completed, only the stage-2 dataset is remained open and available, and it includes 5,244,234 head CT images: one image per one subject. Among 5,244,234 images (subjects), 4,516,842 images are for training, and 727,392 images are for testing. Among the training dataset, 36,118 images are with intraparenchymal, 26,205 images are with intraventricular, 35,075 images are with subarachnoid, 47,166 images are with subdural, 3,145 images are with epidural, and 4,260,600 images are with normal subjects. Note that 7% of ICH patient images include

978-1-7281-1990-8/20/$31.00 ©2020 IEEE

Figure 1. Five type image examples of ICH in the training datset



multiple subtypes. Fig. 1 shows that five different subtype image examples of ICH in the training dataset.

In the testing dataset, the information of ICH presence and its subtype has not been disclosed. For the test, the predicted probability for each different possible subtypes should be submitted to the Kaggle competition website: https://www.kaggle.com/c/rsna-intracranial-hemorrhage-detection.

B. Preprocessing

Fig. 2 shows the preprocessing stage. In the first step, we performed three different windowing with the head CT image. The CT image is composed of a range of CT numbers referred to as the window width (WW). The center of the range is the window level (WL). The process of changing the WW and WL is referred to as windowing. By performing windowing, we obtained three different images: brain window (WW:80, WL:40), bone window (WW:1800, WL:400) and subdural window (WW:200, WL:80). In the second step, we performed the data augmentation such as rescaling, rotation and flipping especially for the epidural, which samples are very few comparing to other subtypes of ICH. In the third step, we balanced the number of each different subtype of ICH images. Finally, we resized the three images to the size of 299x299x3, which are the input for the training.

C. Model Description

The architecture of our proposed model is summarized in Fig. 3. The model consists of deep CNN and LSTM layers. For the deep CNN, we used the Xception model as a backbone, which includes 36 convolutional layers forming the feature extraction base of the network. The 36 convolutional layers are structured into 14 modules, all of which have linear residual connections around them, except for the first and last modules. More specifically, the input data with the size of 299x299x3 first goes through the entry flow, then through the middle flow which is repeated eight times, and finally through the exit flow. All Convolution and separable convolution layers are followed by batch normalization. Subsequently, a LSTM layer with 64 nodes and 32 timesteps was connected to the Xception model via global average pooling layer. Finally, the final output includes 6 nodes with a fully connected layer and sigmoid. In the model, the 6 nodes represent the probabilities of 5 subtypes of ICH (intraparenchymal, intraventricular, subarachnoid, subdural and epidural), and a normal subject.

To avoid overfitting, we applied dropout to the all CNN and LSTM layers. For the convolutional layers, the dropout rates were set to 0.3. For the LSTM layers, the dropout rates for the linear transformation of inputs were 0.3, and the

Figure 2. Data pre-processing

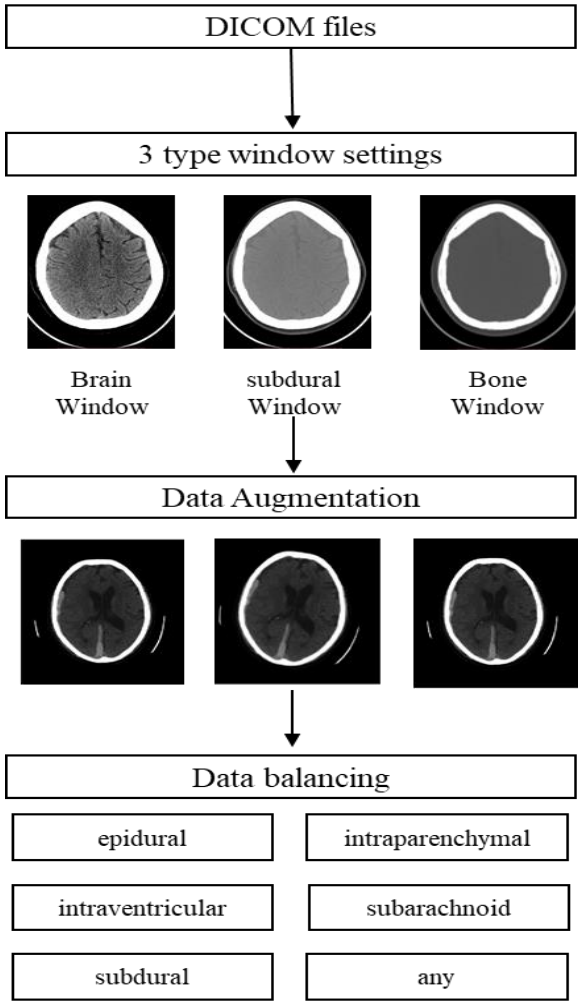
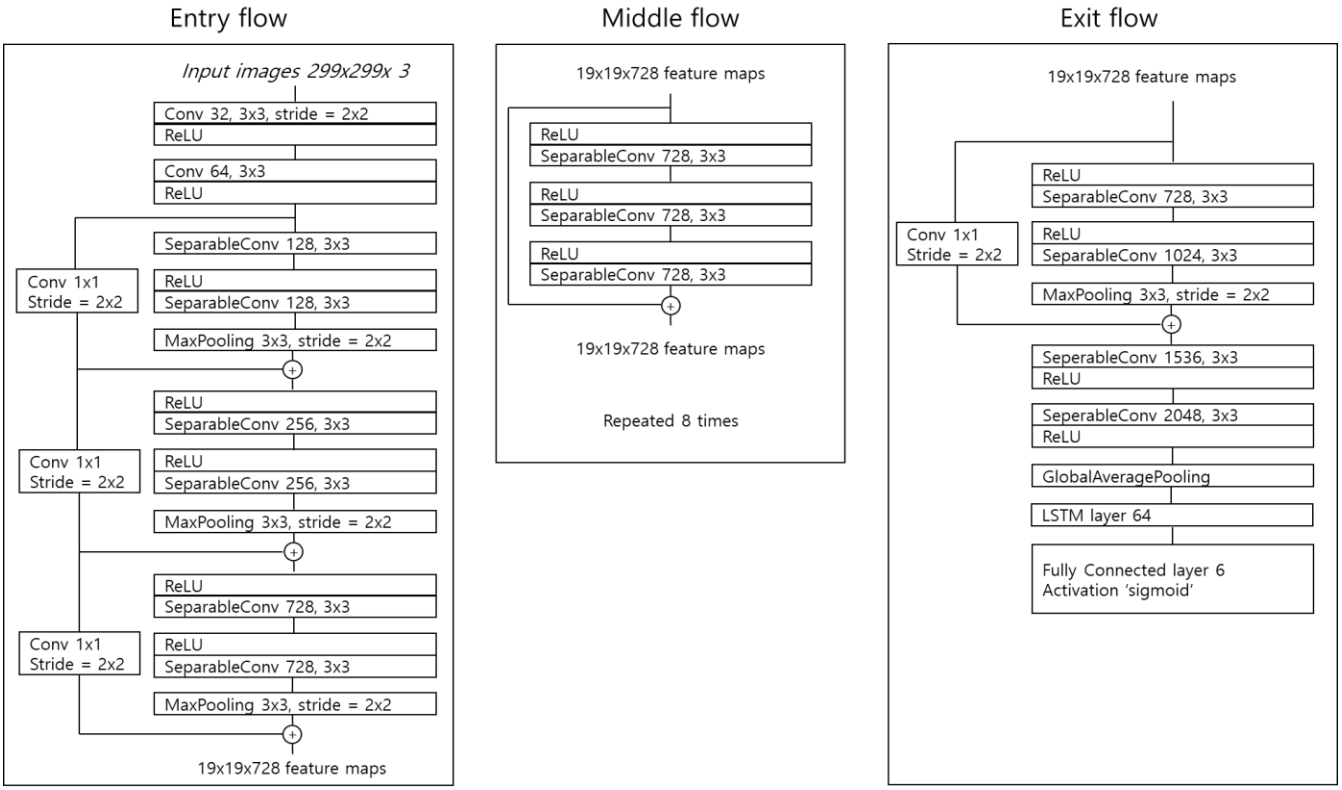


Figure 3. Our model architecture inspired from Xception model



dropout rates for the recurrent state were 0.2. We used multiclass cross-entropy for the cost function.

Our proposed model was implemented using Tensorflow package, which provides a Python API for tensor manipulation for deep learning. We also used Keras, which is now the official frontend of Tensorflow. Keras and Tensorflow, in combination with standard Python libraries such as Numpy and Matplotlib, were used to build the model and analyze the results. We trained our model with the ADAM optimizer with a learning rate of 0.0001 and batch size of 32 on NVIDIA GeForce GTX 1080 Ti GPU.

D. Evaluation Metrics

For each image (subject) in the test set, we predicted a probability for each of the different possible sub-types. Fig. 4 shows the format for the testing.

Given the resultant probabilities, the evaluation was performed using a weighted multi-label logarithmic loss. To avoid the extremes of the log function, the predicted probabilities were replaced with

$$p \leftarrow max(min(p, 1 - 10^{-15}), 10^{-15}) \qquad (1)$$

III. Results and Discussion

For the performance comparison, we trained with two different approaches. The first one was to train with the original unbalanced data. The other one was to train with the balanced data by augmenting each subtype of ICH images. Table I summarizes the comparison results. Based on our proposed model trained with the balanced data, we found that the resultant weighted multi-label logarithmic loss based on the test dataset (N=727,392) was 0.07528, which is approximately equivalent to the classification accuracy of 92 to 93%. With the unbalanced data, we found that the resultant weighted multi-label logarithmic loss based on the test dataset was 0.0813, which is approximately equivalent to the classification accuracy of 88 to 89%. Note since the detail information of ICH presence and its subtype has not been disclosed, that the only way to find the weighted multi-label logarithmic loss results was to submit the results to Kaggle competition website.

Figure 4. Predicted probability for each of the different possible sub-types are with ID and label.

```
Id,Label
1_epidural,0
1_intraparenchymal,0
1_intraventricular,0
1_subarachnoid,0.6
1_subdural,0
1_any,0.9
2_epidural,0
etc.
```

Future work will focus on more extensive clinical testing. The work should include the in-depth analysis on each subtype of ICH. In addition, we should consider and train different deep CNNs such as Inception and VGG for finding optimal model. A variant of LSTM networks such as Gated Recurrent Units (GRU), bi-directional or stacked LSTM also should be considered.

TABLE I. TEST RESULTS WITH COMPARISON BETWEEN BALANCED AND UNBALANCED DATA

Loss	Test result				
	Epoch 1	***Epoch 2***	***Epoch 3***	***Epoch 4***	***Epoch 5***
Unbalanced Data	0.08656	0.08179	**0.0813**	0.08492	0.08942
Balanced Data	0.09882	0.10919	0.0853	0.07776	**0.07582**

IV. CONCLUSION

We have presented the feasibility of the automatic identification of ICH and classification its subtype such as intraparenchymal, intraventricular, subarachnoid, subdural and epidural. We first performed windowing to provide three different images: brain window (WW:80, WL:40), bone window (WW:1800, WL:400) and subdural window (WW:200, WL:80), and trained 4,516,842 head CT images using CNN-LSTM model. We used the Xception model for the deep CNN, and used LSTM with 64 nodes and 32 timesteps. We believe that our proposed method enhances the accuracy of ICH identification and classification and can assist radiologists in the interpretation of head CT images, particularly for brain-related quantitative analysis.

REFERENCES

[1] S. J. An, T. J. Kim, and B.-W. Yoon, "Epidemiology, risk factors, and clinical features of intracerebral hemorrhage: an update," *Journal of stroke,* vol. 19, no. 1, pp. 3, 2017.

[2] J. C. Hemphill, D. C. Bonovich, L. Besmertis, G. T. Manley, and S. C. Johnston, "The ICH score," *Stroke,* vol. 32, no. 4, pp. 891-7, 2001.

[3] V. L. Feigin, C. M. Lawes, D. A. Bennett, S. L. Barker-Collo, and V. Parag, "Worldwide stroke incidence and early case fatality reported in 56 population-based studies: a systematic review," *The Lancet Neurology,* vol. 8, no. 4, pp. 355-369, 2009.

[4] S. Sacco, C. Marini, D. Toni, L. Olivieri, and A. Carolei, "Incidence and 10-year survival of intracerebral hemorrhage in a population-based registry," *Stroke,* vol. 40, no. 2, pp. 394-399, 2009.

[5] C. J. van Asch, M. J. Luitse, G. J. Rinkel, I. van der Tweel, A. Algra, and C. J. Klijn, "Incidence, case fatality, and functional outcome of intracerebral haemorrhage over time, according to age, sex, and ethnic origin: a systematic review and meta-analysis," *The Lancet Neurology,* vol. 9, no. 2, pp. 167-176, 2010.

[6] M. I. Aguilar, and W. D. Freeman, "Spontaneous intracerebral hemorrhage." pp. 555-564.

[7] J. Broderick, S. Connolly, E. Feldmann, D. Hanley, C. Kase, D. Krieger, M. Mayberg, L. Morgenstern, C. S. Ogilvy, and P. Vespa, "Guidelines for the Management of Spontaneous Intracerebral Hemorrhage in Adults: 2007 Update: A Guideline From the American Heart Association/American Stroke Association Stroke Council, High Blood Pressure Research Council, and the Quality of Care and Outcomes in Research Interdisciplinary Working Group: The American Academy of Neurology affirms the value of this guideline as an educational tool for neurologists," *Stroke,* vol. 38, no. 6, pp. 2001-2023, 2007.

[8] J. A. Caceres, and J. N. Goldstein, "Intracranial hemorrhage," *Emergency medicine clinics of North America,* vol. 30, no. 3, pp. 771, 2012.

[9] T. Vedin, S. Svensson, M. Edelhamre, M. Karlsson, M. Bergenheim, and P.-A. Larsson, "Management of mild traumatic brain injury–trauma energy level and medical history as possible predictors for intracranial hemorrhage," *European Journal of Trauma and Emergency Surgery,* vol. 45, no. 5, pp. 901-907, 2019.

[10] I. C. Hostettler, C. Muroi, J. K. Richter, J. Schmid, M. C. Neidert, M. Seule, O. Boss, A. Pangalu, M. R. Germans, and E. Keller, "Decision tree analysis in subarachnoid hemorrhage: prediction of outcome parameters during the course of aneurysmal subarachnoid hemorrhage using decision tree analysis," *Journal of neurosurgery,* vol. 129, no. 6, pp. 1499-1510, 2018.

[11] M. R. Arbabshirani, B. K. Fornwalt, G. J. Mongelluzzo, J. D. Suever, B. D. Geise, A. A. Patel, and G. J. Moore, "Advanced machine learning in action: identification of intracranial hemorrhage on computed tomography scans of the head with clinical workflow integration," *NPJ digital medicine,* vol. 1, no. 1, pp. 1-7, 2018.

[12] J.-E. Scholtz, J. L. Wichmann, D. W. Bennett, D. Leithner, R. W. Bauer, T. J. Vogl, and B. Bodelle, "Detecting intracranial hemorrhage using automatic tube current modulation with advanced modeled iterative reconstruction in unenhanced head single-and dual-energy dual-source CT," *American Journal of Roentgenology,* vol. 208, no. 5, pp. 1089-1096, 2017.

[13] Y.-H. Li, L. Zhang, Q.-M. Hu, H.-W. Li, F.-C. Jia, and J.-H. Wu, "Automatic subarachnoid space segmentation and hemorrhage detection in clinical head CT scans," *International journal of computer assisted radiology and surgery,* vol. 7, no. 4, pp. 507-516, 2012.

[14] Y. Li, J. Wu, H. Li, D. Li, X. Du, Z. Chen, F. Jia, and Q. Hu, "Automatic detection of the existence of subarachnoid hemorrhage from clinical CT images," *Journal of medical systems,* vol. 36, no. 3, pp. 1259-1270, 2012.

[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks." pp. 1097-1105.

[16] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation." pp. 580-587.

[17] F. Chollet, "Xception: Deep learning with depthwise separable convolutions." pp. 1251-1258.