A Deep Learning System for Diagnosing Ischemic Stroke by Applying Adaptive Transfer Learning

Su-Min Jung[1], Taeg-Keun Whangbo[2]

[1] Depart of Computer Science Graduate School of IT, Gachon University, South Korea
[2] Depart of Computer Science, Gachon University, South Korea
sumini1019@nate.com, tkwhangbo@gachon.ac.kr

Abstract

A stroke is the most common, very dangerous single-organ disease and aggravates social burden in the aging society. The stroke can be tested through a variety of imaging methods, among which a test method using CT imaging is known to deal promptly with an emergency patient in the early stage of stroke. Diagnosing ischemic stroke using CT images has advantages such as fewer spatial constrains and quick shooting time. However, diagnosis through images is very difficult, which is a major disadvantage of this method. This study proposed a deep learning system that can conduct learning and classification for ischemic stroke, which is a small dataset and hard to conduct image data learning. This study also proposed a pre-processing algorithm optimized for ischemic stroke based on the non-contrast CT data from the middle cerebral artery (MCA) region. Additionally, this study suggested adopting the adaptive transfer learning algorithm that optimizes the transfer learning module to overcome the problem of insufficient data while training neural networks. When stroke was diagnosed using the proposed system, the performance of it was 18.72% better than the existing system.

Keywords: Stroke, Transfer learning, Deep learning, Brain CT

1 Introduction

Stroke is the most common single-organ disease which claims 6.2 million lives globally each year. Stroke occurs above the age of 65 at a rapid pace, therefore, advanced countries where population ageing is taking place are gradually taking more social burden due to stroke [1-2]. Moreover, this disease is likely to happen commonly in people in their 30s and 40s, therefore, it occurs extensively almost in all age groups and is considered very dangerous. Stroke is divided into an ischemic stroke that occurs when an artery in the brain becomes blocked and cerebral hemorrhage which is caused when an artery in the brain bursts. Figure 1 shows a case in which ischemic stroke and hemorrhagic stroke are evident in non-contrast CT imaging.

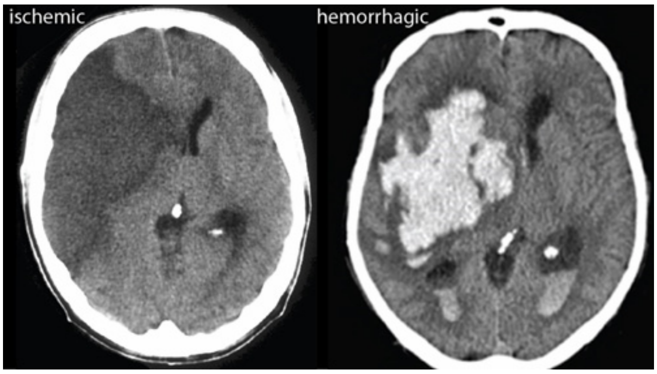


Figure 1. Examples of ischemic-hemorrhagic stroke in brain non-contrast CT

A variety of tests used to diagnose stroke - such as clinical diagnosis, CT (Computed Tomography), MRI (Magnetic Resonance Imaging), and Catheter Angiography - have been developed, of all these, the CT scan has an advantage of bringing fast results. This testing method is regarded as suitable for the characteristics of stroke that requires a prompt response following the occurrence of the disease [3-4]. In particular, now that the use of CT can help determine whether a stroke patient has suffered cerebral hemorrhage, a decision can be made as to whether to use a clot buster provided to patients with cerebral hemorrhage. On top of this, CT is also significantly used in observing the development of stroke following the use of the clot buster.

Studies conducted indicate that when carrying out a test using CT as shown above, excluding the area where a cerebral infarction has rapidly developed may increase the effectiveness of stroke diagnosis and treatment [5-6]. This is implemented with more detail in so-called ASPECTS (Alberta Stroke Program Early CT Score) [7-8].

But, the determination of early signs of ischemia and their translation into the ASPECTS have a considerable inter-rater variability, which is, among other factors,

*Corresponding Author: Taeg-Keun Whangbo; E-mail: tkwhangbo@gachon.ac.kr
DOI: 10.3966/160792642020122107010

influenced by rater's experience [9]. Hence, inter-rater variability depending on rater's experience has a very negative impact on decision making about the patient's stroke identification. One solution to improve ASPECTS readings is to train doctors to be aware of these issues and provide strategies that enhance the reliability and validity of these readings. Another solution is to develop an automated solution for interpreting ASPECTS using new technologies such as machine learning and feature extraction.

This study developed a CAD system that could automatically classify strokes based on machine learning and image processing techniques. Two neurosurgeons diagnosed stoker using the same NCCT data to compare the performance by comparing it with the results of a deep learning system. The stroke ground-truth (label) was generated by conducting cross-examination and feedback.

2 Materials and Methods

2.1 Problems and Limitations of Previous Studies

This study proposed a classification neural network model that was effective for diagnosing stroke by analyzing and improving the problems of the existing studies on stroke. The main problems of the existing stroke research are as follows.

(1) Lack of medical image data

It is hard to construct a large medical image dataset for a specific disease because of issues such as patients' privacy protection and the independent data management of each hospital [10]. An insufficient dataset causes an overfitting problem. Moreover, the estimated weight of a neural network is not sufficient to show sufficient performance. Therefore, it is difficult to accurately classify disease through a deep learning system.

(2) Voxel CT data learning

Brain NCCT imaging continuously scans from the top of the patient's head at approximately 25mm slice thickness, including the MCA region. It produces 2D slices (40-50) per patient, and each image is obtained in the form of a Dicom. Consequently, it has metadata information such as the coordinate values for the x, y, and z axes. The 2D image data of the same patient is serial data, and it is possible to know the sequence information based on the metadata of the corresponding coordinate values [11]. However, since a CNN model using a 2D single image has a problem of not learning the relationship between continuous images, a neural network system considering Voxel CT data is required.

(3) Performance limitations of a single CNN model

This study aimed to classify and diagnose stroke diseases by training the neural network. It is very difficult to construct a large dataset due to problems such as insufficient previous data [12]. When the volume of a dataset is small like this case and model learning is conducted only with the given training data, it may cause problems such as impossible to learn feature and overfitting. The performance of the neural network shall be improved by applying an algorithm such as data augmentation to overcome these problems. These algorithms are methods for improving performance in individual models, and performance improvement in individual models has a clear limitation. Consequently, it is necessary to improve the performance of the entire system by creating a model that is more generalized and can avoid overfitting, which can be achieved by diversifying training and prediction neural network models based on the same dataset and combining training weights and predictions from multiple models.

(4) Medical image classification deep learning model feedback problem

Large-capacity public data, such as ImageNet and CIFAR-10, is composed of images that contain information about common subjects, such as an object that can be easily understood by the general public. For these images, a system developer can determine the label of each image and annotate the label even if each data does not contain a label.

Most commercial deep learning systems perform learning and prediction based on supervised learning, and this process requires a large dataset. Therefore, most AI companies and multinational companies invest a lot of manpower and capital in securing data and constructing label data for the data [13-14].

On the other hand, it is impossible to diagnose a disease using imaging data and to compose label data unless someone has professional knowledge in diagnostic radiology. Due to this issue, it is greatly difficult to construct label data for medical imaging [15-16]. Therefore, it is hard to use a system utilizing object detection and semantic segmentation models, which require a lot of time and labor to construct label data among deep learning classification models for images. Consequently, it mainly uses simple classification models that are relatively simple to prepare label data. However, in the classification model, it is difficult to find out what region of the input image is used to generate the prediction result [17-18]. Considering that the objective of the deep learning system application to medical imaging is to explore special regions in images and diagnose, verification and feedback of neural network's prediction results are essential and it is possible to improve the performance of the entire system through this.

2.2 Overview of the Suggested Algorithm

This study proposed a deep learning CAD system that provides a basis for judgment prior to diagnosis by medical personnel. The proposed system classifies the stroke of patients based on image processing and deep learning technology using NCCT images obtained from patients with suspected stroke symptoms and suggests appropriate medical treatments according to the patient's stroke type and stage.

The proposed deep learning-based stroke classification CAD system aims to resolve the problems and limitations described in Section 2.1 by using the following methods.

It was induced to detect and learn the cerebral infarction that occurred locally only in a part of the brain through the concatenation of sequential slices to take into account the continuous 2D data of CT images for the MCA region.

This study pre-processed the data entered into the neural network by optimizing the windowing algorithm, which converts data from HounsField Units(HU), CT raw data, by acquiring data distribution for each stroke type through analyzing the histogram of ischemic stroke patients' cerebral infarction regions. It improves the performance of neural network learning for ischemic stroke's cerebral infarction region.

The optimal initial weight was explored through the transfer learning algorithm to overcome the lack of ischemic stroke data. It made a small amount of ischemic stroke data converge to the optimal performance. The source data required for transfer learning was based on publicly available hemorrhagic stroke data similar to ischemic stroke. Moreover, this study proposed adaptive transfer learning to overcome the increase in the complexity of transfer learning and operation quantity according to the depth of the neural network.

The deep learning-based stroke classification system consists of four stages: (1) hemorrhagic stroke data pre-processing; (2) hemorrhagic stroke classification neural network; (3) ischemic stroke data pre-processing; and (4) ischemic stroke classification neural network. Figure 2 depicts the overall process of the deep learning-based stroke classification system.

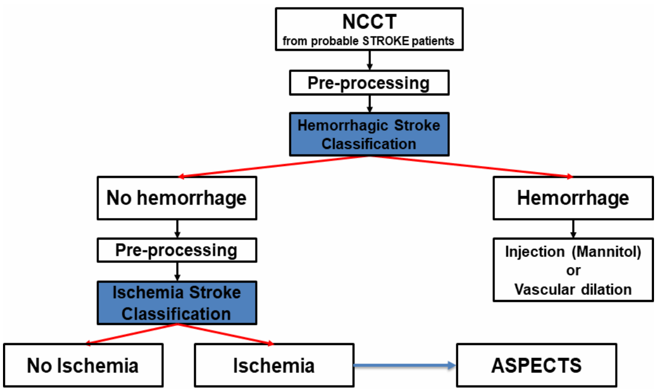


Figure 2. A structure of deep learning-based stroke classification CAD system

The system quantifies the degree of ischemic stroke by performing primary classification for hemorrhagic stroke based on NCCT images, diagnosing the presence of the disease in the ischemic stroke classification neural network after carrying out an ischemic stroke specialized pre-process for diagnosing ischemic stroke for patients whose hemorrhagic stroke is negative in the hemorrhagic stroke classification neural network, and calculating the ASPECT score based on the presence of cerebral infarction for each brain region.

It is much more difficult to construct the ischemic stroke data due to insufficient data compared to hemorrhagic stroke. Therefore, this study mainly tried to improve this issue and this study derived the best performance by optimizing the pre-processing and neural network structure by using the previously proposed improvement algorithm. Figure 3 shows the schematic structure of the ischemic stroke classification system.

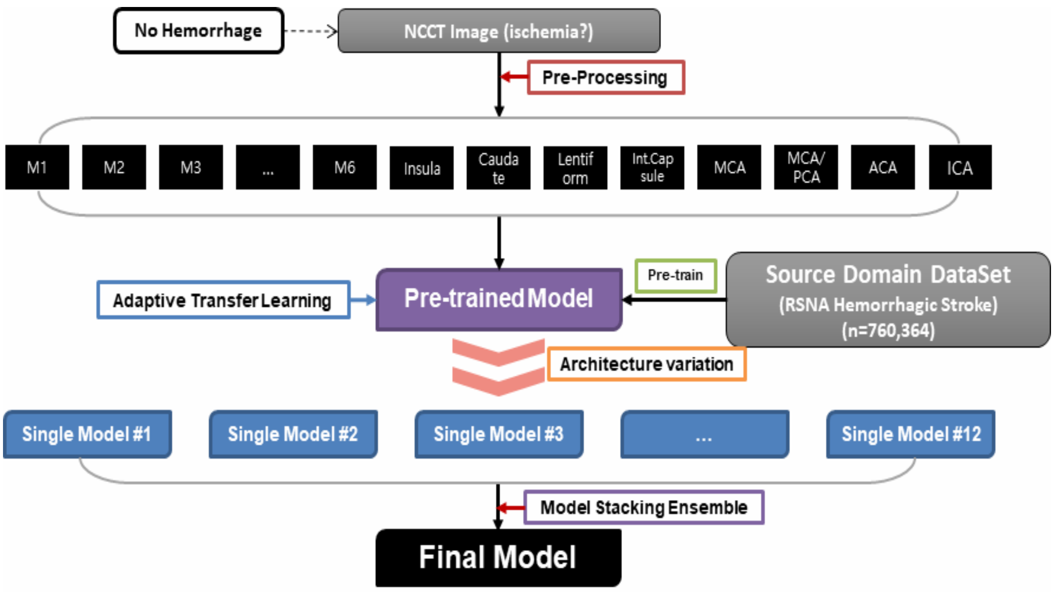


Figure 3. A structure of ischemic stroke classification CAD system

2.3 Lesion Concatenation

Data pre-processing includes the template conversion of CT raw data for normalizing inter-datasets, data augmentation, lesion concatenation between multi slices, and windowing. The proposed system constructs data with multiple slices per patient by reconstructing and normalizing brain NCCT images. The neural network trains the model by receiving each target slice data and classifies whether it is a stroke or not. Although the target slice entered into the neural network is based on a 2D image, the actual brain NCCT image is composed of sequence data from the top of the head to the MCA region.

Moreover, the proposed system segments each region of the bran and classifies the cerebral infarction of each region. A specific region is composed of three sequential slices, up to three slices, based on a slice thickness of 3mm. Therefore, when infarction occurs in at least one slice out of multiple slices, the neural network must determine the region of a patient as cerebral infarction. Figure 4 shows an example of old infarction occurring only in the last slice among three sequential slices in the insula region. Since infarction occurred in slice #3, although infarction did not occur in slices #1 and #2, the region should be classified as old infarction.

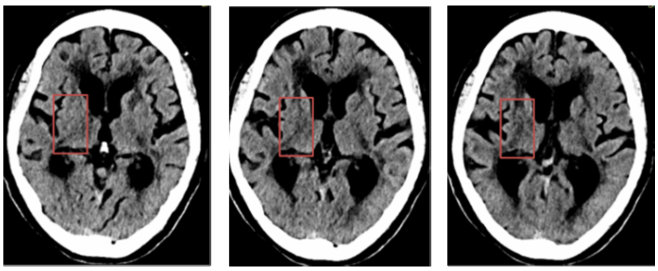


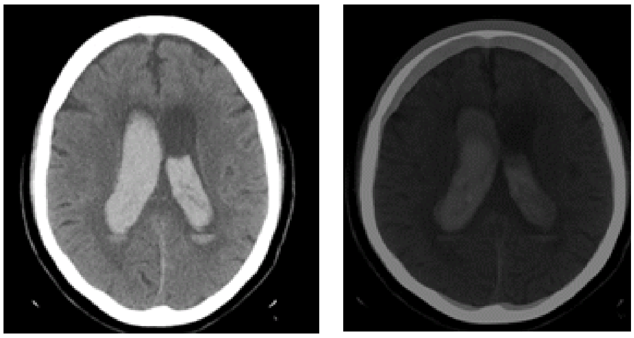
Figure 4. Examples of infarction occurrence in sequential slices of a region

This study explores the before-after slices of the target slice based on the relative position meta-data and enters multiple sequential slice images for one region into the neural system to apply this clinical know-how for stroke diagnosis to the deep learning system. The neural network performs concatenation for the entered sequential slices and converts it into one image to perform model learning and prediction.

Figure 5 shows a single target slice and the data that applies lesion concatenation to a single target slice. The image to which lesion concatenation is applied is calculated using the before and after images of the target slice.

2.4 Optimal Windowing

CT data has a CT numerical value for each pixel, which is called HounsField Units (HU) after the name of Godfrey N. Hounsfield, a UK inventor who developed the CT technique. The value presents the relative absorption by the density of each part when X-rays penetrate the body: 0 for water, 1000 for bone, and -1000 for air. The range of HU is from –1000 to 3000, a wide range. Diagnostic radiology does not consider the full range of HU and restructures it to a certain range of HU to visualize a certain region. This process is called windowing or window leveling. The windowing method displays only the value of a specific window width around the HU value of interest (window center). The values outside the window width, among HU values, are converted to 0 or 255, and the values within the window width are expressed as a distribution ranging from 0 to 255. The conversion formula is usually a linear function and Recale slope and intercept are often used as the parameters of this. It can be formulated like the following equation (1).



(a) a single target slice and (b) a lesion concatenated slice

Figure 5. Slices before and after lesion concatenation

$$\begin{cases} \textbf{\textit{Lowest Value}} = WindowCenter - (\frac{WindowWidth}{2}) \\ \textbf{\textit{Lowest Value}} = WindowCenter - (\frac{WindowWidth}{2}) \\ \textbf{\textit{Value}} = (HU * \mathrm{Re}\,scaleSlope) + RescaleIntercept \end{cases} \quad \textbf{(1)}$$

In the case of ischemic stroke, the HU distribution of the normal group and that of the infarction group are little different. Figure 6 presents the mean histogram of each class of data. It shows that the distribution of the data is similar and the difference between the mean values is very small. The results indicate that it is very important to limit the range of HU data by applying appropriate windowing.

This study proposes optimal windowing parameters that can classify the early ischemic sign class (Frank Hypodensity, Territorial Infarction), which is the main target infarction of ischemic stroke, with high performance, and each value is window width 20-30 and window center 40. Figure 7 shows the histogram of the left and right brain HU. One side is an infarction corresponding to the early ischemic sign class. When the opposite brain is normal, the difference between the HU can be found in a certain pattern when the difference between the two histograms is calculated. By setting each parameter based on the pattern, the

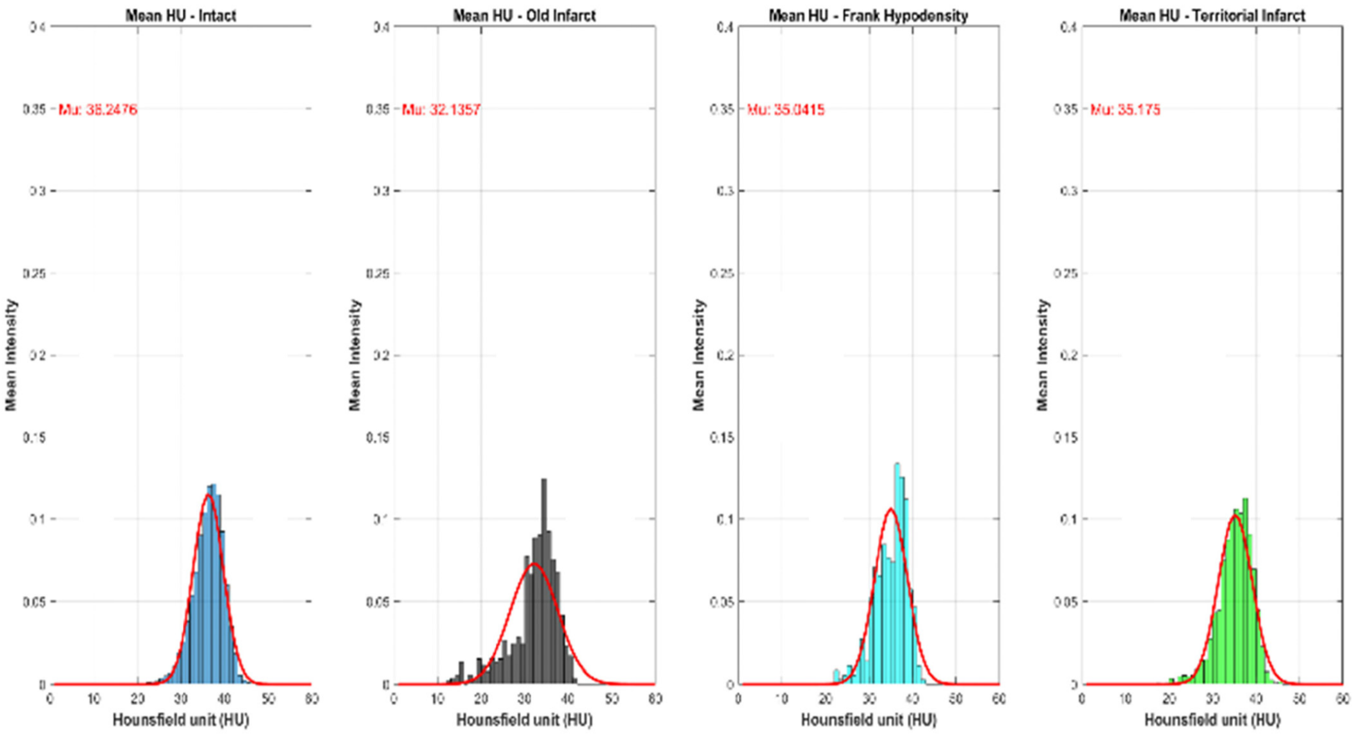


Figure 6. Mean histograms of datasets by normal and infarction

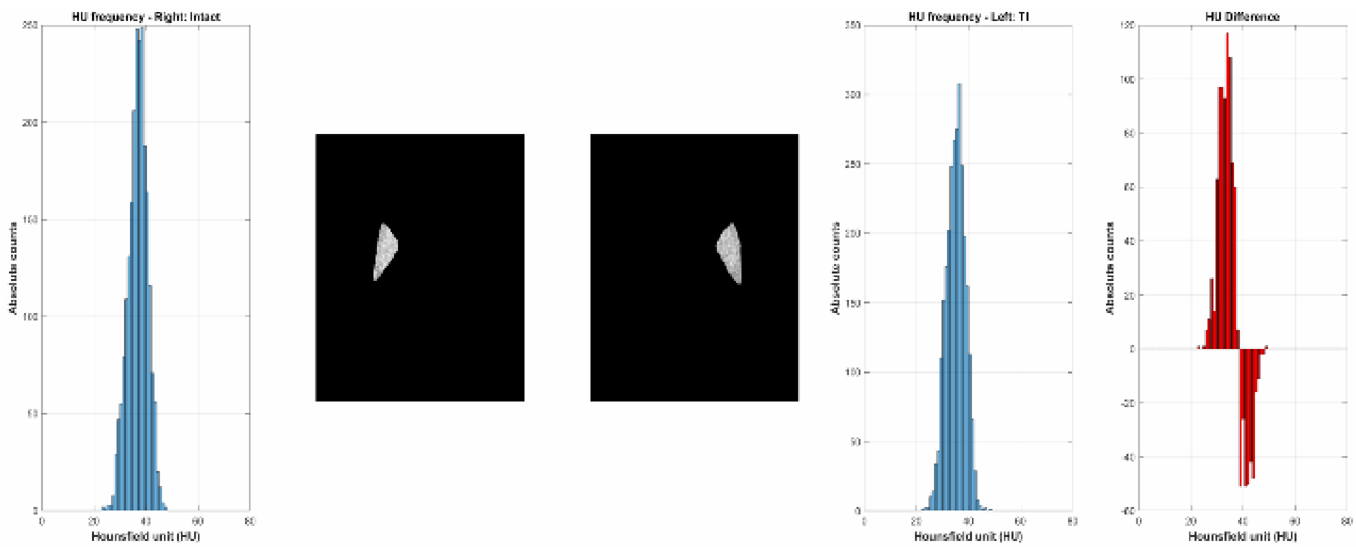


Figure 7. Histograms in the infarction and normal ranges and the difference between them

image is restructured by only using HU values between 25 and 55 within the HU range of -1,000 and 3,000. When it goes through the proposed windowing, optimal learning can be expected by using only pixel information corresponding to the features of ischemic stroke in the image.

2.5 Adaptive Transfer Learning

The biggest problems in classifying ischemic stroke are the small differences in features between the normal group and the infarction group and insufficient data volume. Among them, this study tried to overcome the insufficient data volume issue using data augmentation. The proposed study intended to improve the performance of neural network learning through transfer learning as well as data manipulation such as direct data volume increase. Transfer learning is a concept of learning the model ex-post by transferring a deep learning model learned for a specific source task to a target task. In general, when applying transfer learning for learning a specific target task, it can show better performance with a smaller amount of learning data than simply using a randomly initialized weighted deep learning model to learn that task.

Figure 8 presents traditional machine learning, and different tasks train independent models based on each data.

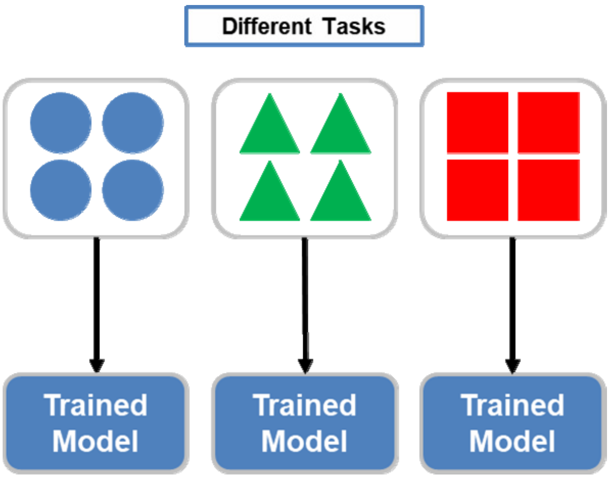


Figure 8. Traditional machine learning training

Transfer learning generates weights optimized for the source task data by learning a model with data corresponding to the source task. The knowledge of the source task is utilized later when learning a deep learning model by using the data of the target task to be solved. Figure 9 presents the concept of transfer learning.

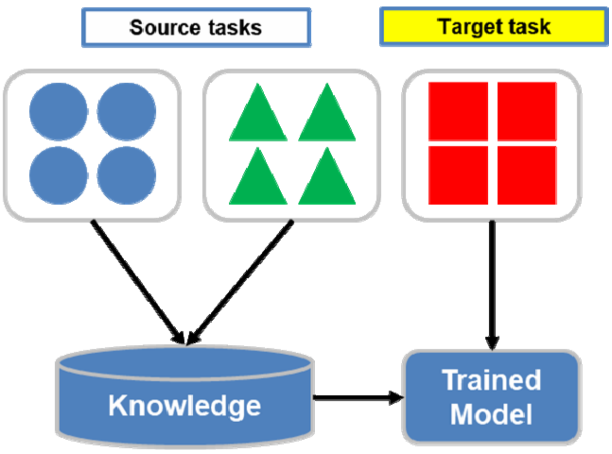


Figure 9. Conceptual diagram of transfer learning

Since the transfer learning conducts learning based on the weight of the existing source task, as shown above, it has higher initial and final performance expectations. Moreover, it can overcome the insufficient data issue of the target task and has the advantage of improving the speed of learning and optimization. However, transfer learning does not always guarantee good performance, and factors determining the performance of transfer learning must be optimized. Factors determining the performance of transfer learning include the similarity of a source task and a target task, the volume and consistency of task datasets, layer freeze due to task characteristics and neural network structure, and a fine-tuning strategy [19]. Figure 10 shows the transfer learning strategy theory according to the size of a source task and the similarity between tasks.

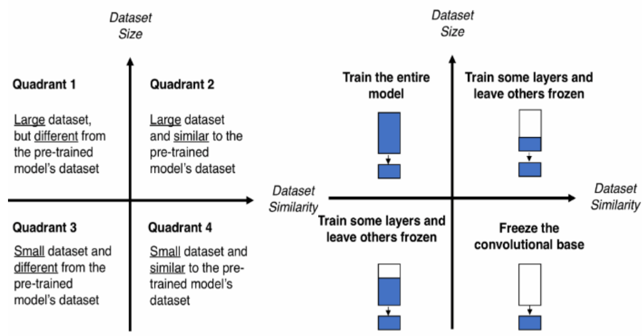


Figure 10. Transfer Learning Strategy

Transfer learning freezes the neural network layer shallower – starting from the part near to the output layer toward the input layer- and conducts fine tuning for weight about unfrozen layer when the size of a source task is larger and the similarity between tasks is smaller. Fine tuning of the weight is performed based on the data of the target task, and the frozen layer does not perform the weight update through fine tuning.

The problem of transfer learning is to require a lot of time and computer resources to explore layer freeze-fine tuning that yields optimal performance while applying the transfer learning strategy. The transfer learning strategy is determined by the amount of the source task dataset and the similarity between tasks. These parameters cannot be quantified and it is based on the subjective judgment of experts. Moreover, the ResNet neural network used as the top model in ILSVRC2015 consists of 152 layers. In such a complex neural network, it is impossible to determine the optimal layer freezing depth, and it is necessary to compare and analyze each performance while adjusting the depth. Figure 11 shows the problem of determining the optimal layer freeze depth when applying transfer learning to ResNet-152 neural networks with 152 layers.

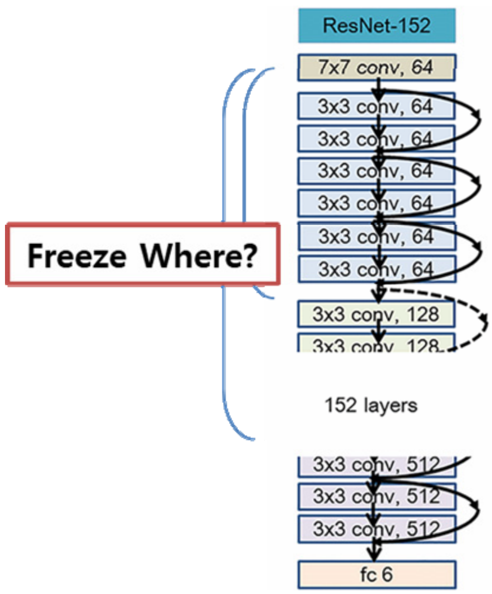


Figure 11. Problems in optimizing transfer learning for ResNet-152

Therefore, this study proposed adaptive transfer learning that optimizes transfer learning so that can explore the target task neural network model that shows optimal performance by inputting the target task dataset and the neural network model of each task.

The proposed adaptive transfer learning, while conducting transfer learning, automatically explores the optimal layer freezing depth, proposes a transfer learning performance index according to the layer freezing depth, and proposes an automated algorithm based on this. The transfer learning performance index is calculated based on binary cross entropy losses, and optimization is performed by calculating a performance index for each transfer learning step. The conceptual diagram of the proposed adaptive transfer learning is shown in Figure 12.

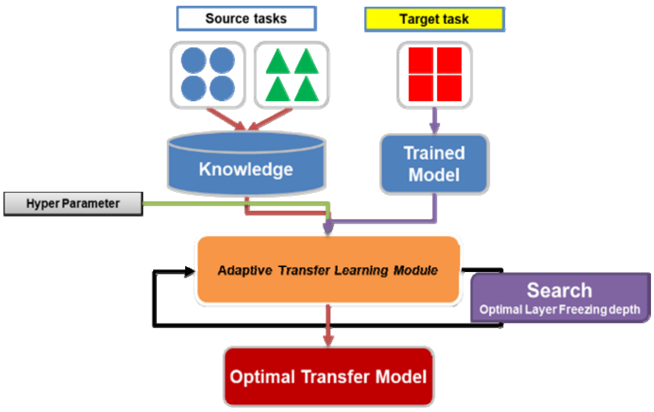


Figure 12. A conceptual diagram of adaptive transfer learning

It calculates a transfer learning model based on the optimized by exploring the optimal layer freezing depth after applying the adaptive transfer learning algorithm described in equation (2) by receiving the knowledge model learned through the data of the source task and the data and model of the target task. The algorithm sequentially freezes the layer from the input layer of the neural network to the output layer and calculates the binary cross entropy loss (BCE Loss) of each step. The function ***f*** is defined by adding the initial BCE Loss value and the minimum BCE Loss value of each step, and the variable $\boldsymbol{\theta}$ that minimizes the function ***f*** is calculated.

$$\begin{cases} When, \begin{cases} \theta : Layer\ Freezing\ Depth, \\ t : Target\ Task, \\ w_{1,2} : Weight\ of\ Loss \\ L_{s\to t} : Loss\ of\ Model \\ \qquad \begin{pmatrix} Transfered\ by\ Source\ Task \\ Fine\ Tuned\ by\ get\ Task \end{pmatrix} \end{cases} \\ \boldsymbol{argmin(f(\theta)) = w_1 L_{s\to t_{frz=0,e_1}} + \frac{w_2(\sum_{k=1}^{e} L_{s\to t_{frz=0,e_k}})}{e}} \end{cases} \quad (2)$$

3 Results and Discussion

3.1 Experimental Environment

An AMD Ryzen9 3900X 12-Core Processor desktop was used for the experiment, and the operating system was Windows 10 Home. One Geforce RTX 2080 Ti was used as a GPU. Python 3.7 was used for implementing the algorithm, and Python-OpenCV 3.0 was used for the development of the pre-processing algorithm. Pytorch 1.4.0 was used for learning and testing the deep learning algorithm.

This study utilized the data of 356 stroke patients, which was accumulated over 8 years in a hospital to evaluate the algorithm proposed in this study. The data of each patient was composed of 40-50 pieces of 2D images according to the slice thickness, when CT is taken. Eight pieces per patient were used among 2D images reconstructed through template normalization. Moreover, in the eight 2D images, 46 2D images were obtained by dividing them by the side of the brain for 14 regions of an ASPECTS target. Learning and testing were carried out using 16,376 2D images: 13,064 images were used for learning, and 3,312 images were used for testing. All images were entered into the neural network after they were converted into 512 × 512 images. The batch size was fixed at 16. The test was performed by selecting a model showing the lowest cross entropy loss value after performing epoch 100 times.

Each image had ground-truth data and it was recorded in the form of a csv file. Ground-truth was recorded in one of six categories for the 14 regions of each patient: intact, old infarction, recent infarction, frank hypodensity, territorial infarction, and scattered infarction

3.2 Lesion Concatenation Application and Performance Evaluation

This study learned and tested a neural network by applying inter-image concatenation to effectively classify cerebral infarction by using before-after slice information for the region composed of sequential slices among cerebral infarction candidate regions. The performance of each learned neural network was compared through the neural network, which learned only single slices without using lesion concatenation, and the data that was applied by lesion concatenation for sequential slices (Table 1). The neural network evaluated the classification performance for the intact group and the old infarction group and the classification performance for the intact group and the early ischemic sign group [20].

Table 1. Performance evaluation before and after applying lesion concatenation

Evaluation Index	Case #1	Case #2	Case #3	Case #4
Number of Train Image	5,210	5,210	8,234	8,234
Precision	0.7784	0.7841	0.6244	0.6201
Recall (Sensitivity)	0.7390	0.7407	0.5740	0.5679
F-1 Score	0.8088	0.8260	0.6883	0.6619
AP*	0.6810	0.6994	0.5363	0.5318
Accuracy	0.8048	0.8269	0.6891	0.6857
AUC**	0.8592	0.8634	0.7195	0.7098

Note. *: Average Precision; **: Area Under the Curve (ROC Curve); *Case #1*: Without Lesion Concatenation (Classify Old Infarction); *Case #2*: With Lesion Concatenation (Classify Old Infarction); *Case #3*: Without Lesion Concatenation (Classify Early Ischemic Sign); *Case #4*: With Lesion Concatenation (Classify Early Ischemic Sign).

Old infarction is a class that is easier to classify because it has a more pronounced difference in HU. When lesion concatenation was applied, it showed a 2.02% performance increase. However, the early ischemic sign group, which was hard to be classified due to a small difference with an intact group, did not show significant performance change owing to lesion concatenation.

3.3 Optimal Windowing Application and Performance Evaluation

The images of the intact group and that of the cerebral infarction class have very little difference in the distribution of HU, CT raw data. Therefore, the performance of feature learning of the neural network is determined according to preprocessing that limits the range of HU values. This transfer algorithm is called windowing. This study explored the optimized parameters by setting windowing parameters suitable for learning neural network features based on the HU distribution by the intact group and cerebral infarction classification and comparing the neural network performance according to it.

Table 2 shows the classification performance for the intact group and the old infarction group according to the windowing parameter setting. Table 3 presents the results of evaluating the classification performance for the intact group and the early ischemic sign group.

Table 2. Performance evaluation according to windowing parameters (Old infarction classification)

Evaluation Index	WC: 40, WW: 40 (Classify Old Infarction)	WC: 40, WW: 30 (Classify Old Infarction)	WC: 40, WW: 20 (Classify Old Infarction)
Number of Train Image	5,210	5,210	5,210
Precision	0.7784	0.7924	0.6426
Recall (Sensitivity)	0.7390	0.7497	0.6071
F-1 Score	0.8088	0.8308	0.7098
AP*	0.6810	0.7073	0.5900
Accuracy	0.8048	0.8307	0.6912
AUC**	0.8592	0.8691	0.7207

Note. *: Average Precision; **: Area Under the Curve (ROC Curve).

Table 3. Performance evaluation according to windowing parameters (Early ischemic sign classification)

Evaluation Index	WC: 40, WW: 40 (Classify Early Ischemic Infarction)	WC: 40, WW: 30 (Classify Early Ischemic Infarction)	WC: 40, WW: 20 (Classify Early Ischemic Infarction)
Number of Train Image	8,234	8,234	8,234
Precision	0.6244	0.6047	0.6381
Recall (Sensitivity)	0.5740	0.5843	0.6445
F-1 Score	0.6883	0.6892	0.7198
AP*	0.5363	0.5380	0.5706
Accuracy	0.6891	0.6921	0.7013
AUC**	0.7195	0.7141	0.7401

Note. *: Average Precision; **: Area Under the Curve (ROC Curve).

The classification of the intact group and the old infarction group showed the highest performance under Window Center 40 and Window Width 30, when the windowing range was [25, 55]. On the other hand, the classification of the intact group and the early ischemic sign group under Window Center 40 and Window Width 30, when the windowing range was [30, 50]. The HU distribution of old infarction had characteristics of a lower mean (4.11 lower) and standard deviation than that of the intact group. Moreover, that of the early ischemic sign group had approximately 1.74 lower values and a similar standard deviation compared to that of the intact group. Therefore, it is possible to confirm the possibility of learning the features composing the old infarction only after performing windowing with including a wider range of HU compared to the classification of the early ischemic sign group.

Therefore, when classifying ischemic stroke, it is necessary to apply individual windowing parameters according to the type of stroke. This study proposed windowing parameters optimized for each cerebral infarction classification. They are Window Center 40 and Window Width 30 when classifying old infarction and Window Center 40 and Window Width 20 when classifying early ischemic sign.

3.4 Adaptive Transfer Learning Application and Performance Evaluation

This study overcame the performance issue due to the insufficient dataset by conducting transfer learning based on the hemorrhagic stroke dataset of the Radiological Society of North America, which had relatively large data and something in common that the data was the NCCT imaging for the brain MCA region.

This study proposed adaptive transfer learning to overcome the problems of neural network layer freezing, fine tuning depth determination strategy selection, and transfer learning velocity. Moreover, this study verified the superiority of the proposed algorithm by comparing before and after the proposed adaptive transfer learning application.

In the experiment, the adaptive transfer learning calculated the performance indicators by automating the layer freezing and fine tuning of the SEResNext neural network model to calculate performance comparison indicators. Table 4 shows the classification performance of the intact group and the early ischemic sign group by the freezing step. A neural network model (not transfer) that learned using only ischemic stroke data without adaptive transfer learning was set as the control group. The control group used the SEResNext neural network model as in the experimental group. The control group's neural network trained ischemic stroke data using the weights pre-trained with ImageNet as initial values.

Table 4. Performance comparison by adaptive transfer learning step

Evaluation Index	Not Freeze	Freeze Layer #0	Freeze Layer #0~1	Freeze Layer #0~2	Freeze Layer #0~3	Freeze Layer #0~4	Freeze All Conv-Pooling Layer	Not Transfer
Number of Train Image	8,234	8,234	8,234	8,234	8,234	8,234	8,234	8,234
Precision	0.6364	0.6667	0.6244	0.5794	0.5778	0.4799	0.6244	0.6267
Recall (Sensitivity)	0.6278	0.6278	0.5740	0.6547	0.6996	0.8027	0.5740	0.6323
F-1 Score	0.6963	0.7110	0.6745	0.6606	0.6614	0.5489	0.6745	0.6883
AP*	0.5549	0.5740	0.5363	0.5235	0.5297	0.4676	0.5363	0.5498
Accuracy	0.6948	0.7135	0.6742	0.6554	0.6610	0.5674	0.6742	0.6891
AUC**	0.7176	0.7374	0.6901	0.6995	0.7166	0.6538	0.6901	0.7195

Note. *: Average Precision; **: Area Under the Curve (ROC Curve).

A model with a deepened neural network layer freezing depth (Freeze Layer #0~4, Freeze All Conv-Pooling Layer) did not learn normally because it either did not conduct learning for the ischemic stroke data, the target task, or learned only a very few features. Figure 13 shows the predicted value of each image when predicting a test dataset using the "Freeze Layer #0-4" learning neural network. Although it was a binary classification, it only generated prediction values between 0.4 and 0.65, indicating that normal learning was not performed.

When the neural network layer freezing (reducing the fine-tuning layer) was deepened, the recall tended to increase. In all performance indicators except the recall, the model, which froze up to the layer #0 corresponding to the front end of the neural network layer and fine-tuned other layers through the ischemic stroke, showed the highest performance. Figure 14 shows the ROC Curve and Area Under the Curve of the model frozen up to Layer #0.

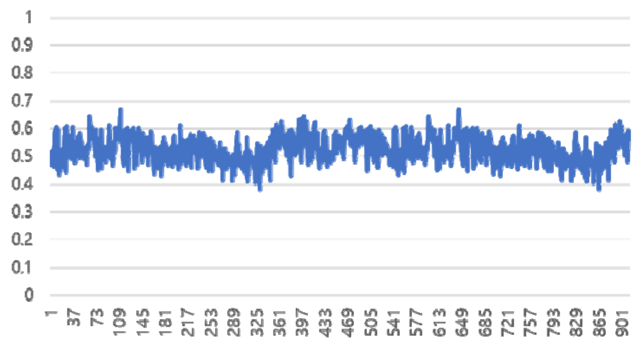


Figure 13. The distribution of predicted value with deep neural network layer freezing (Freeze Layer #0~4 – SeResNext Model)

Figure 15 compares performance indicators according to the neural network freezing through adaptive transfer learning and fine-tuning depth.

The neural network model optimized through adaptive transfer learning improved performance by approximately 3.43% compared to the model without it.

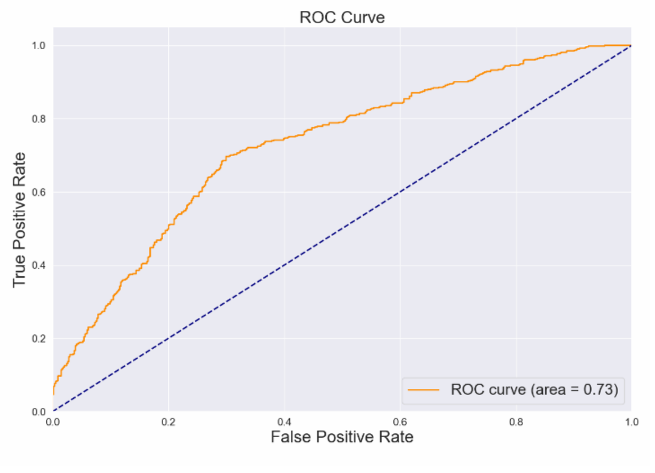


Figure 14. ROC Curve and Area Under the Curve of the model frozen up to Layer #0

It can be concluded that, compared to a single training model using ischemic stroke data, conducting adaptive transfer learning by using hemorrhagic stroke data as a source task significantly improved the performance.

Moreover, when performing general transfer learning, a source task and a target task are similar, and the target task has relatively little data, fine tuning only the neural network layers at the back end and freezing many layers are accepted as a general strategy. However, the results of the experiment showed the opposite result. This proves that the strategy selection of transfer learning according to unquantifiable values is very lacking and it is necessary to directly compare the performance according to the neural network layer freezing through adaptive transfer learning.

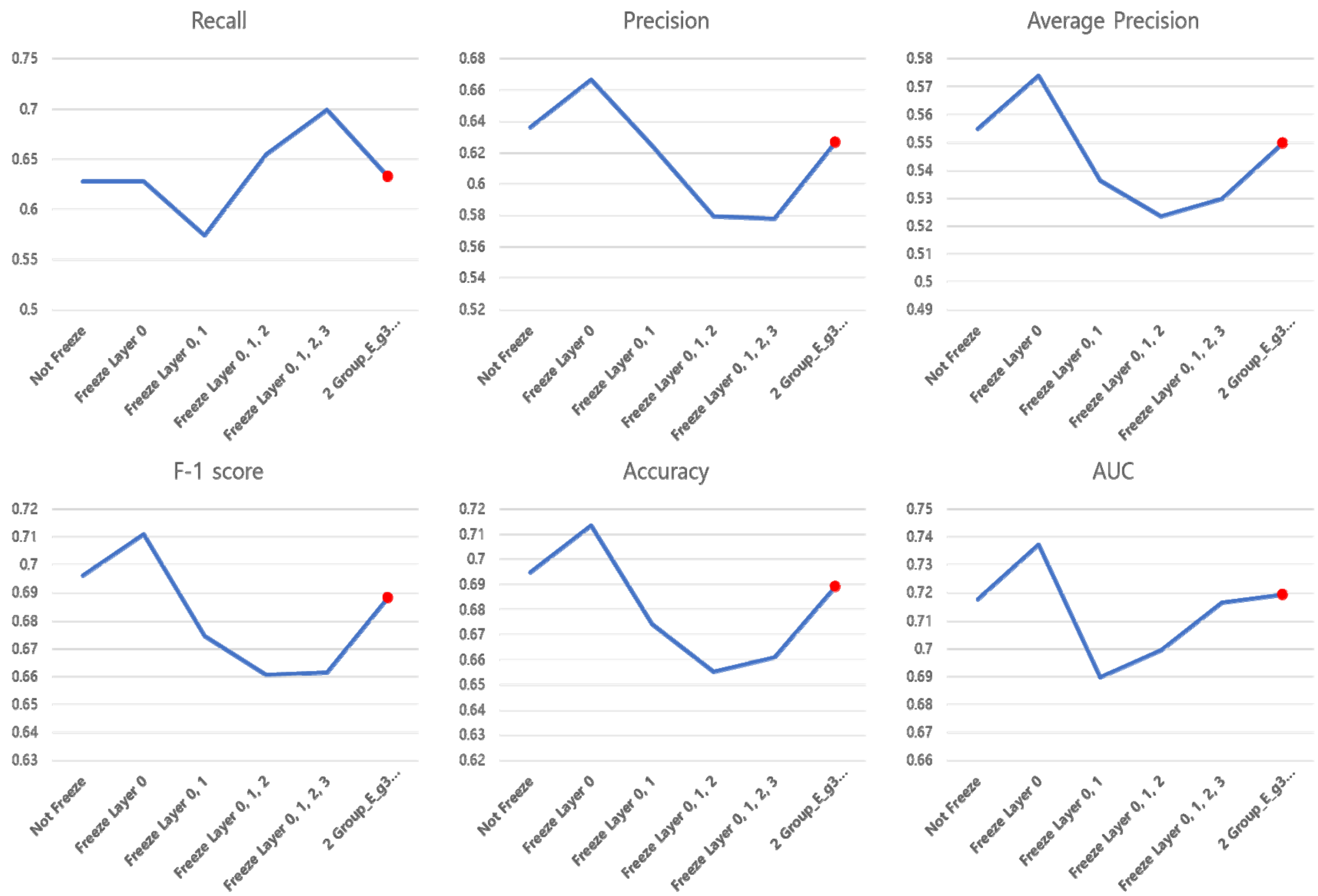


Figure 15. Performance indicators by adaptive transfer learning step

4 Conclusion

The importance of medical imaging recognition technology based on artificial intelligence is increasing due to the aging of the population and the lack of chronic medical personnel. Along with the development of deep learning technology, many studies have been conducted on various diseases and the accuracy of it has also improved to a considerable level.

However, medical images suffer from an insufficient data issue for establishing a reliable deep learning system because they have a large variation in data according to imaging devices, an issue of patients' privacy protection, and the independent data management issue of each medical service provider. Moreover, the deep learning system utilizing CT and X-ray medical images has been studied much less than the MRI imaging-based deep learning system providing highly reliable diverse information.

Additionally, it is difficult to secure performance for marketability [21]. Therefore, there is a high demand for a deep learning system based on X-ray and CT images, which can be taken quickly, for diseases that require rapid medical treatment such as stroke.

This study proposed a deep learning-based stroke classification model using non-contrast CT medical images to overcome these problems. This study conducted neural network learning through lesion concatenation to consider the serial data of CT images in the form of a voxel. Furthermore, this study presented windowing parameters optimal for detecting ischemic cerebral infarction. Additionally, to overcome the issue of insufficient ischemic stroke data, this study tried to achieve the best performance using a small amount of ischemic stroke data by exploring the optimal initial weight through the adaptive transfer learning algorithm. The source data required for the adaptive transfer learning algorithm was performed based on public data of hemorrhagic stroke, which is similar to ischemic stroke, and transfer learning optimization was carried out by resolving the transfer learning complexity and computation amount increase issues through the proposed adaptive transfer learning.

The proposed method trained and evaluated the presence and type of stroke by using non-contrast CT images of 356 stroke patients, obtained for 8 years in the medical field. The experimental data were classified into intact, old infarction, and early ischemic sign (frank hypodensity, territorial infarction) classes to apply and evaluate each proposed algorithm. The performance was compared with the performance of the existing SEResNext neural network, which did not apply the proposed algorithm under the same condition. Based on the data which was pre-processed by applying the proposed algorithm and the improved neural network system, cerebral infarction CT data was learned and classified. The results showed that the performance improved by approximately 18.72% compared to the existing neural network learning and test on the original NCCT images. The performance evaluation produced four performance indicators (average precision, F-1 score, accuracy, and the AUC of the ROC curve) by determining the class with the highest prediction value as the prediction class based on the prediction value of a cerebral infarction class.

This study proposed a deep learning-based automated stroke classification system that could calculate an aspect score, an objective indicator for diagnosing the condition of a stroke patient, using only CT images. It is expected that the stroke classification program can be used as a reliable indicator that can prevent the issue of inter-expert scoring variability issue and can aid medical personnel to make medical decisions easier considering the nature of stroke disease requiring rapid treatment. The system proposed by this study can not only aid people to determine the initial treatment through ischemic stroke diagnosis but also help the initial response to cerebral diseases, for which first aid is critical, because it can be used as a parameter in judging emergent large vessel occlusion (ELVO) to decide the application of a thrombolytic agent to an acute stroke patient by obtaining ASPECTS, a quantified score. Moreover, the system can be actively used to observe the prognosis of stroke by tracking changes in the cerebral infarction of the patient over time after treatment as well as the diagnosis of the initial ischemic stroke. It is expected that it can be used as an index for additional treatment and medication prescription.

Future studies are needed to improve the current system classification, which classifies only the presence of stroke in the medical image, to a semantic segmentation form neural network. When the system is improved to a semantic segmentation neural network, it will be possible to overcome the performance degradation of neural network learning according to the pre-processing accuracy, which is a shortfall caused by replacing the ROI process for the region in the medical image with the brain region segmentation pre-processing. Moreover, it is believed that it will be able to secure the marketability because it can have a clinical advantage of clear lesion detection.

It is also expected that all modules can be utilized as a neural network learning improvement module by utilizing limited data in the medical imaging deep learning field, including the pre-processing module for the NCCT data in the arteriae mesencephalic region and neural network performance improvement through adaptive transfer learning, which were tested and evaluated in this study.

Acknowledgements

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2020-2017-0-01630) supervised by the IITP (Institute for Information) & Communications Technology Promotion.

References

[1] S. Mendis, P. Puska, B. Norrving, *Global Atlas on Cardiovascular Disease Prevention and Control*, World Health Organization, 2011.

[2] G. A. Roth, D. Abate, K. H. Abate, S. M. Abay, C. Abbafati, N. Abbasi, I. Abdollahpour, Global, Regional, and National Age-sex-specific Mortality for 282 Causes of Death in 195 Countries and Territories, 1980-2017: A Systematic Analysis for the Global Burden of Disease Study 2017, *The Lancet*, Vol. 392, No. 10159, pp. 1736-1788, November, 2018.

[3] J. Aoki, K. Kimura, K. Shibazaki, Y. Sakamoto, DWI-ASPECTS as a Predictor of Dramatic Recovery after Intravenous Recombinant Tissue Plasminogen Activator

Administration in Patients with Middle Cerebral Artery Occlusion, *Stroke*, Vol. 44, No. 2, pp. 534-537, February, 2013.

[4] T. Nezu, M. Koga, K. Kimura, Y. Shiokawa, J. Nakagawara, E. Furui, H. Yamagami, Y. Okada, Y. Hasegawa, K. Kario, S. Okuda, K. Nishiyama, M. Naganuma, K. Minematsu, K. Toyoda, Pretreatment ASPECTS on DWI predicts 3-month Outcome Following rt-PA: SAMURAI rt-PA Registry, *Neurology*, Vol. 75, No. 6, pp. 555-561, August, 2010.

[5] R. Von Kummer, K. L. Allen, R. Holle, L. Bozzao, S. Bastianello, C. Manelfe, E. Bluhmki, P. Ringleb, D. H. Meier, W. Hacke, Acute Stroke: Usefulness of Early CT Findings before Thrombolytic Therapy, *Radiology*, Vol. 205, No. 2, pp. 327-333, November, 1997.

[6] R. Von Kummer, H. Bourquain, S. Bastianello, L. Bozzao, C. Manelfe, D. Meier, W. Hacke, Early Prediction of Irreversible Brain Damage after Ischemic Stroke at CT, *Radiology*, Vol. 219, No. 1, pp. 95-100, April, 2001.

[7] J. H. W. Pexman, P. A. Barber, M. D. Hill, R. J. Sevick, A. M. Demchuk, M. E. Hudon, W. Y. Hu, A. M. Buchan, Use of the Alberta Stroke Program Early CT Score (ASPECTS) for Assessing CT Scans in Patients with Acute Stroke, *American Journal of Neuroradiology*, Vol. 22, No. 8, pp. 1534-1542, September, 2001.

[8] P. A. Barber, A. M. Demchuk, J. Zhang, A. M. Buchan, ASPECTS Study Group, Validity and Reliability of a Quantitative Computed Tomography Score in Predicting Outcome of Hyperacute Stroke before Thrombolytic Therapy, *The Lancet*, Vol. 355, No. 9216, pp. 1670-1674, May, 2000.

[9] J. Pfaff, C. Herweh, S. Schieber, S. Schönenberger, J. Bösel, P. A. Ringleb, M. Möhlenbruch, M. Bendszus, S. Nagel, e-ASPECTS Correlates with and Is Predictive of Outcome after Mechanical Thrombectomy, *American Journal of Neuroradiology*, Vol. 38, No. 8, pp. 1594-1599, August, 2017.

[10] H. C. Park, Y. J. Kim, S. W. Lee, Adenocarcinoma Recognition in Endoscopy Images Using Optimized Convolutional Neural Networks, *Applied Sciences*, Vol. 10, No. 5, Article Number: 1650, March, 2020.

[11] J. Lim, B. Wang, J. S. Lim, A Hierarchical Two-phase Framework for Selecting Genes in Cancer Datasets with a Neuro-fuzzy System, *Technology and Health Care*, Vol. 24, No. s2, pp. S601-S605, June, 2016.

[12] R. K. Lama, S. W. Lee, White Matter Network Alterations in Alzheimer’s Disease Patients, *Applied Sciences*, Vol. 10, No. 3, Article Number: 919, February, 2020.

[13] J. Y. Li, J. H Li, Prompt Image Search with Deep Convolutional Neural Network via Efficient Hashing Code and Addictive Latent Semantic Layer, *Journal of Internet Technology*, Vol. 19, No. 3, pp. 949-957, May, 2018.

[14] N. Q. Nguyen, D. M. Vo, S. W. Lee, Contour-aware Polyp Segmentation in Colonoscopy Images Using Detailed Upsamling Encoder-Decoder Networks, *IEEE Access*, Vol. 8, pp. 99495-99508, May, 2020.

[15] D. Wang, J. Tian, T. K. Whangbo, Method for Real-time Automatic Setting of Ultrasonic Image Parameters Based on Deep Learning, *Multimedia Tools and Applications*, Vol. 78, No. 1, pp. 1067-1080, January, 2019.

[16] S. Poudel, Y. J. Kim, D. M. Vo, S. W. Lee, Colorectal Disease Classification using Efficiently Scaled Dilation in Convolutional Neural Network, *IEEE Access*, Vol. 8, pp. 99227-99238, May, 2020.

[17] N. Q. Nguyen, S. W. Lee, Robust Boundary Segmentation in Medical Images Using a Consecutive Deep Encoder-decoder Network, *IEEE Access*, Vol. 7, pp. 33795-33808, March, 2019.

[18] D. M. Vo, N. Q. Nguyen, S. W. Lee, Classification of Breast Cancer Histology Images Using Incremental Boosting Convolution Networks, *Information Sciences*, Vol. 482, pp. 123-138, May, 2019.

[19] Q. Cui, Z. Zhou, C. Yuan, X. Sun, Q. J. Wu, Fast American Sign Language Image Recognition Using CNNs with Fine-tuning, *Journal of Internet Technology*, Vol. 19, No. 7, pp. 2207-2214, December, 2018.

[20] M. Zhou, Z. Bai, T. Yi, X. Chen, W. Wei, Performance Predict Method Based on Neural Architecture Search, *Journal of Internet Technology*, Vol. 21, No. 2, pp. 385-392, March, 2020.

[21] R. K. Lama, J. Gwak, J. S. Park, S. W. Lee, Diagnosis of Alzheimer’s Disease Based on Structural MRI Images Using a Regularized Extreme Learning Machine and PCA Features, *Journal of Healthcare Engineering*, Vol. 2017, Article ID 5485080, June, 2017.

Biographies

Su-Min Jung received a bachelor's degree in 2013 from Gachon University, Korea, and a master’s degree in 2015 from Gachon University, Korea. Mr. Jung has been working as an algorithm developer for one year in a vision and image processing company in Korea. He is currently pursuing a Ph.D. in computer engineering from Gachon University, Korea. His research areas include image processing, deep learning, healthcare service, and AR/VR/MR.

Taeg-Keun Whangbo received a M.S. degree in computer science from City University of New York, USA in 1988 and a Ph.D. in computer science from Stevens Institute of Technology, USA in 1995. Currently, Mr. Whangbo is a professor in the Department of Interactive Media, Gachon University, Korea. He is also the Dean of Research Affairs at Gachon University. Before he joined Gachon University, Mr. Whangbo was a software developer at Q-Systems, Inc., New Jersey, USA from 1988 to 1993. He was also a researcher at Samsung Electronics, Korea from 2005 to 2007. From 2006 to 2008, Mr. Whangbo was the president of the Association of Korea Cultural Technology. His research areas include computer graphics, HCI, and AR/VR/MR.