Journal of Population Therapeutics & Clinical Pharmacology

RESEARCH ARTICLE
DOI: 10.47750/jptcp.2023.30.06.042

A Study on the Classification of Cancers with Lung Cancer Pathological Images Using Deep Neural Networks and Self-Attention Structures

Seung Hyun Kim[1], Ho Chul Kang[2*]

[1]Student, Department of Medical Artificial Intelligence, The Catholic University of Korea, Korea

[2]Professor, Department of Media Technology and Media Contents, The Catholic University of Korea, Korea

*Corresponding author: Ho Chul Kang, Professor, Department of Media Technology and Media Contents, The Catholic University of Korea, Korea, Email: hckang19@catholic.ac.kr

Submitted: 17 February 2023; Accepted: 10 March 2023; Published: 09 April 2023

ABSTRACT

In this paper, we propose a ResNet-based lung cancer pathology image classification model using deep neural networks and self-attention modules. With a shortcut structure that adds input as an output, which is ResNet's concept, we not only solve the vanishing gradient problem but also perform well even when layers are piled densely. Based on this idea, the pre-activation structure in which the output enters the input as it is was used by moving the position of batch normalization and activation function in front of the weight layer was used. ResNet's bottleneck structure is made up of layers with 1x1, 3x3, and 1x1 convolution layers, which are utilized for depth wise convolution and 1x1 convolution, respectively, to conduct convolution operations in the channel direction. In addition, channel attention and spatial attention were used as self-attention modules that focus on certain features after the bottleneck structure. Finally, batch normalization and activation functions were used, and after using the Funnel Activation function considering two-dimensional space as the activation function, the model is constructed with a fully connected layer with average pooling and activation function as sigmoid. The accuracy, precision, recall, and f1-score of our method are 82.83%, 83%, 84.14%, and 83.56 respectively. We show through experiments that our method is better than existing ResNet-based models.

Keywords: *Image Classification, Deep Learning, Self-Attention, Depthwise Convolution, ResNet, Convolution Neural Network, Lung Cancer Classification*

INTRODUCTION

Deep learning is attracting attention by improving performance in various fields such as speech recognition, natural language processing, and computer vision. Deep learning is carrying out machine learning using artificial neural networks with multiple layers. It learns important patterns and rules from large-scale data, and carries out decision-making, prediction, etc. based on the learning.

As deep learning has been applied in the field of medicine, it has been utilized in various fields such as the departments of radiology, ophthalmology, and dermatology, but studies utilizing deep learning are the most frequently conducted in the department of pathology among them. Diagnosis in the department of pathology is carried out by analyzing tissue or cell samples collected from the patient to determine whether any tumor exists or not and whether the tumor is malignant or benign, and predict the patient's prognosis. Unlike general images in which characteristics such as color, shape, and texture commonly appear across images, images used in pathology have different patterns distributed diversely, making it very difficult to analyze them [2][3]. In addition, although diagnoses are carried out after seeing the images through a microscope or digital scanner, the results are different by doctor and deviations occur in the diagnosis results because doctors judge after seeing the images with their eyes [1] [16]. Therefore, deep learning technology is applied to increase diagnostic efficiency and accuracy.

This study proposes employing the Self-attention Module [6][7][8] to focus on characteristics in the pathological images of lung cancer and applying Depthwise Convolution [5] to the structure of the residual block based on ResNet [4] as a cancer classification model. Through experiments, the method proposed in this paper showed higher performance than the existing ResNet.

The structure of this paper is as follows. Chapter 2 describes ResNet [4] and Self-Attention Module [6] [7] [8] as models related to this study, and a new activation function [11] applied to the model. Chapter 3 describes the model proposed in this study. explain Chapter 4 introduces the dataset used for learning and validation, the experimental method and results, and Chapter 5 describes the conclusion of this study. The composition of this paper is as follows. Chapter 2 describes ResNet [4] and Self-Attention Module [6] [7] [8] as models related to this study, and a new activation function [11] applied to the models, and Chapter 3 describes the model proposed in this study. Chapter 4 introduces the dataset used for learning and validation and shows the experimental method and results, and Chapter 5 describes the conclusion of this study.

THEORY

ResNet

In the case of neural network models, as the layers are accumulated deeply, the problems of gradient vanishing and gradient explosion occur, leading to the problem of performance deterioration [4]. Gradient vanishing is a phenomenon in the gradient value becomes very small as the backpropagation process progresses. As a way to solve this problem, ResNet's Residual Block appeared. Below Figure 1. is a figure that shows the existing neural network (Plane Layer) and the Residual Block. The existence or absence of the skip connection distinguishes the two methods. The skip connection refers to the process through which the output value skips the intermediate layer and is added to the input of the next layer. The existing neural network receives input x and outputs H(x) through the layer, aiming at finding function H(x) that maps input x to target value y. In the case of the Residual Block, input x is received as usual and F(x), which is residual information, is additionally added [4].

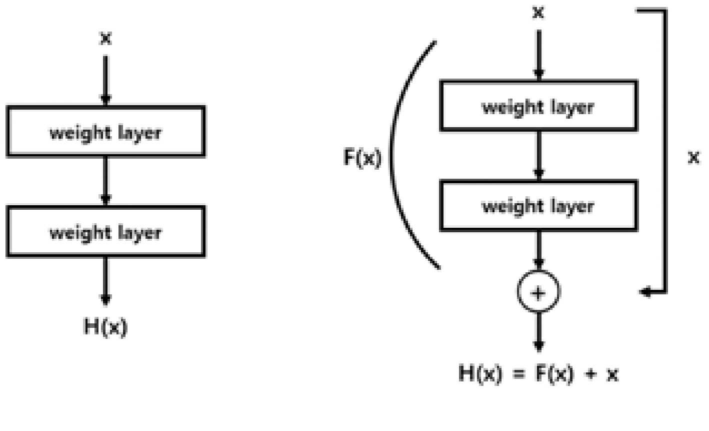


FIGURE 1: Plane Layer & Residual Block

Depthwise Convolution

Since the standard convolution filter is affected by all channels of the input image, it is impossible to extract spatial features of only certain channels. A method proposed to supplement the foregoing problem is the Depthwise Convolution, which uses filters that are implemented only for their respective single channels. That is, convolution in the channel direction is not carried out and only convolution in the spatial direction is carried out. In addition, since the convolution multiplication is independently carried out by the channel, the amount of computation is reduced compared to the standard convolution.

Channel and Spatial Attention

The concept of attention means concentrating on a certain feature and was first applied to the field of natural language processing [7]. Recently, as it has been applied to convolutional neural networks, it has been used in various fields such as Image Captioning, Image Classification, and Object Detection. Thereafter, as the concept termed self-attention, which means the occurrence of attention, was created in some network modules, Channel Attention [7], Spatial Attention [8], etc. have been studied.

Channel Attention emphasizes certain channels using the relationship between the feature map and channels. The structure is shown in Figure 2. Below and is expressed as a formula as shown in (1). First, x, which is the input feature map, becomes a one-dimensional vector as much as the channel size through Global Average Pooling (GAP), and each value becomes a representative value of the corresponding channel. Thereafter, channel weights are generated using a 1x1 convolution layer with a kernel size of k, a vector with an emphasized channel size is created through a nonlinear activation function (σ) sigmoid, and the vector and x are multiplied to create y, an output feature map.

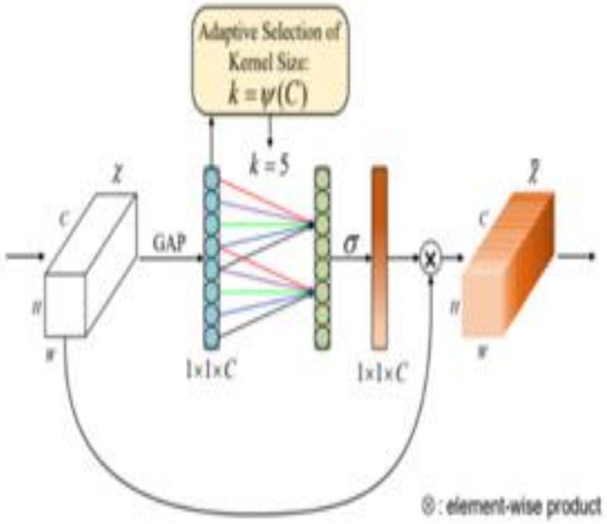


FIGURE 2: Channel Attention Module [7]

$$y = \sigma(conv_{1x1}\big(GAP(x)\big) \cdot x \qquad (1)$$

The structure of Spatial Attention is shown in Figure 3. below, and can be expressed as a formula as shown in (2). A feature map attended to one channel is created by applying input feature map x to a 1x1 convolution layer (conv1x1) to collect channel-by-pixel data. The feature map highlighted using sigmoid, a non-linear activation function (), is then multiplied by x to form an attended feature map (y).

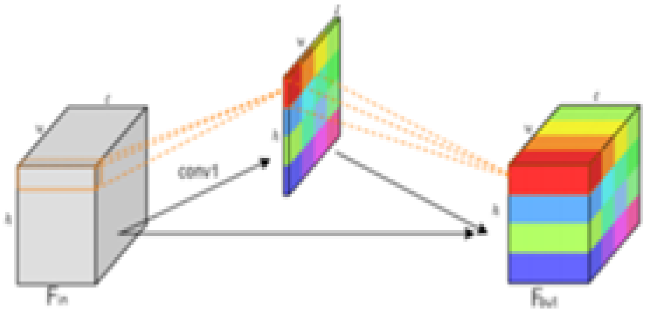


FIGURE 3: Spatial Attention Module [8]

$$y = \sigma(conv_{1x1}(x)) \cdot x \qquad (2)$$

Funnel Activation (FReLU)

Activation functions such as ReLU [9] and PReLU [10], which are widely used in CNNs, have a problem that they do not consider spatial elements, and FReLU was designed considering 2D spaces as a way improve the problem as such. FReLU shows improved performance in image classification, object detection, and semantic segmentation. A figure comparing ReLU, PReLU, and FReLU is Figure 4. below, and the formula representing FReLU is defined as (3). where, $x_{c,i,j}$ means the position (i, j) of the c-th channel of the image, and T(.) refers to a function of the two-dimensional space and is called the Funnel Condition. In the T(.) expression, $x^{w}{}_{c,i,j}$ is called a Parametric Pooling Window, which is a window of kh x kw centering on $x_{c,i,j}$x_(c,i,j), and the window size is basically 3x3. $p^{w}{}_{c}$ is different by channel, identically to the p in the expression $max(x, 0) + p \cdot min(x, 0)$ of PReLu. Eventually, it can be seen that T(.) is identical to a depth wise convolution.

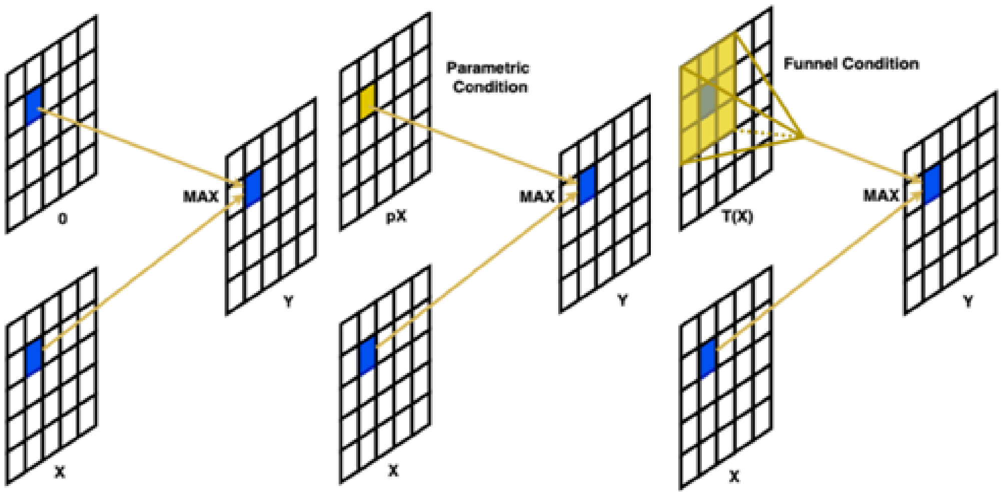


FIGURE 4: ReLu & PReLu & Funnel Activation(FReLu) [11]

$$f(x_{c,i,j}) = max(x_{c,i,j}, T(x_{c,i,j})) \qquad (3)$$

$$T(x_{c,i,j}) = x^{w}{}_{c,i,j} \cdot p^{w}{}_{c}$$

PERFORMANCE IMPROVEMENT

Input Stem Change

An ResNet consists of an Input Stem, four Stages, and an Output Layer. The Input Stem consists of one 7x7 Convolution Layer and a 3x3

Maxpooling Layer, but the 7x7 Convolution Layer has a problem in that the amount of computation greatly increases along with the size of the kernel. Therefore, a method to reduce the amount of computation by processing data with two 3x3 Convolutions was proposed in Inception-V2[12]. In this study, as with the method proposed in Inception-V2, two 4x4 convolutions were used with a smaller kernel size than that of the 7x7 convolution layer to reduce the amount of computation. In addition, since the size of the feature map decreases as the depth of the layer increases, the stride was set to 2 and the number of filters to 32 for the first convolution and the number of filters was set to 64 for the second convolution [13].

Changes in the Downsampling Block

In the structure of ResNet, among four stages, Stage-2 and the following stages consist of one Downsampling Block and one Residual Block. The Downsamplig Block is divided into two paths, Path-A and Path-B. In the case of Path-A, three convolution layers with kernel sizes of 1x1, 3x3, and 1x1 are formed. The stride size of the first convolution layer is 2 to reduce the width and height of the input value by half and the output channel of the last convolution layer is four times that of the previous two convolution layers. This is also called a bottleneck structure. Path-B has a 1x1 Convolution Layer, and the Stride Size is set to 2 to adjust the output size to be equal to that of Path-A to combine the outputs of the two paths. The residual Block is similar to the Downsampling Block except for the fact that it uses a Convolution Layer with a stride size of 1. In this study, the ResNet was divided into a Downsampling Block and a Residual Block from Stage-1, and since there was a problem that a part of the Feature Map disappeared when the Stride Size of the first layer is set to 2 in Path-A of the existing Downsampling Block of the ResNet [13], the stride size of the first convolution layer of Path-A was changed to 1 and the stride size of the second convolution layer was changed to 2 in the Downsampling Block and Residual Block of all stages to solve the problem. To solve the problem of Feature Map loss occurring in Path-B as with Path-A, 2x2 average pooling was added in front and the stride size of the 1x1 convolution layer was changed to 1 [13]. In addition, in ResNet, it is important to wrap the input value with Identity Mapping and deliver it as the output value so that information is not lost because this will prevent the gradient loss problem from occurring. Therefore, Batch Normalization using the idea as such and the Pre-Activation [14] structure in which the position of the activation function was changed were utilized.

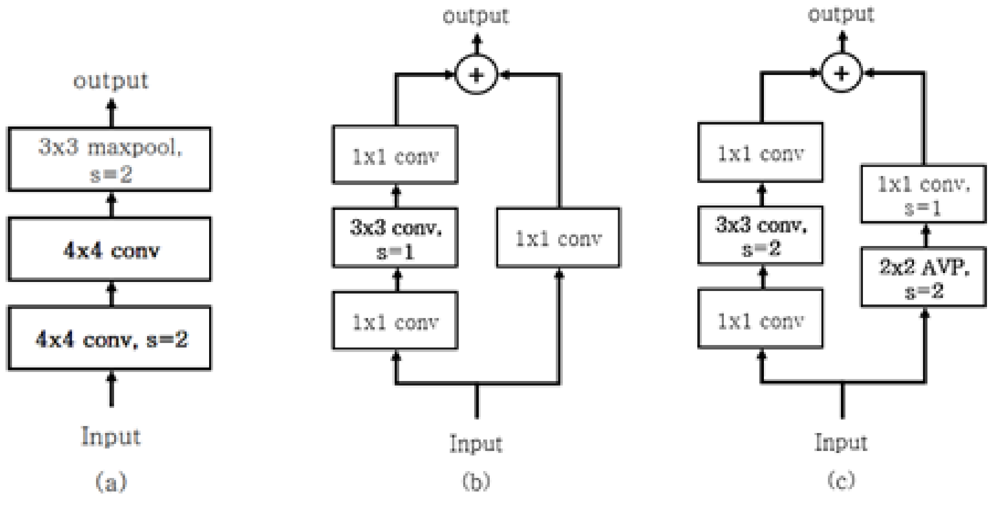


FIGURE 5: The Image for Performance Improvement of (a) input stem, (b) residual block and (c) downsampling block

Application of Convolution and Attention Techniques by Depth

In order to reduce the amount of computation in the first 1x1 convolution layer, the channel was first forcibly shrunk in the Path-A 1x1, 3x3, and 1x1 convolution layer structures. Then, features were extracted through the 3x3 convolution layer, and the channel was again expanded in the final 1x1 convolution layer. The depthwise convolution was applied before the first and last 1x1 convolution layers because the 1x1 convolution layer is the same as the pointwise convolution, which only performs operations for individual channels without extracting the spatial feature for the input. As a result, the convolution played the same role as the Depthwise Separable Convolution [15] to extract channels and spatial features. Second, after the last 1x1 convolution of the bottleneck structure, channel attention and spatial attention were sequentially applied with the attention technique. Then, the channel attention made the input feature map into a one-dimensional vector of 1x1xC in which important information was compressed through GAP, generated channel weights using a 1x1 convolution layer with a kernel size of k, and multiplied the one-dimensional vector that underwent the non-linear activation function by the Input Feature Map before being applied with the GAP to make an Output Feature Map. Thereafter, the Output Feature Map made through channel attention in spatial attention a was applied to a 1x1 convolution with a kernel size of k to collect channel information by pixel to make attended feature maps by channel. Thereafter, a nonlinear activation function sigmoid was used and the Feature Map emphasized through the Sigmoid was multiplied by the abovementioned Output Feature Map finally to extract the characteristic feature map. A fully connected layer and a nonlinear activation function sigmoid were utilized once normalization was completed at each level, applied to FReLU as a nonlinear activation function, and then again through batch normalization.

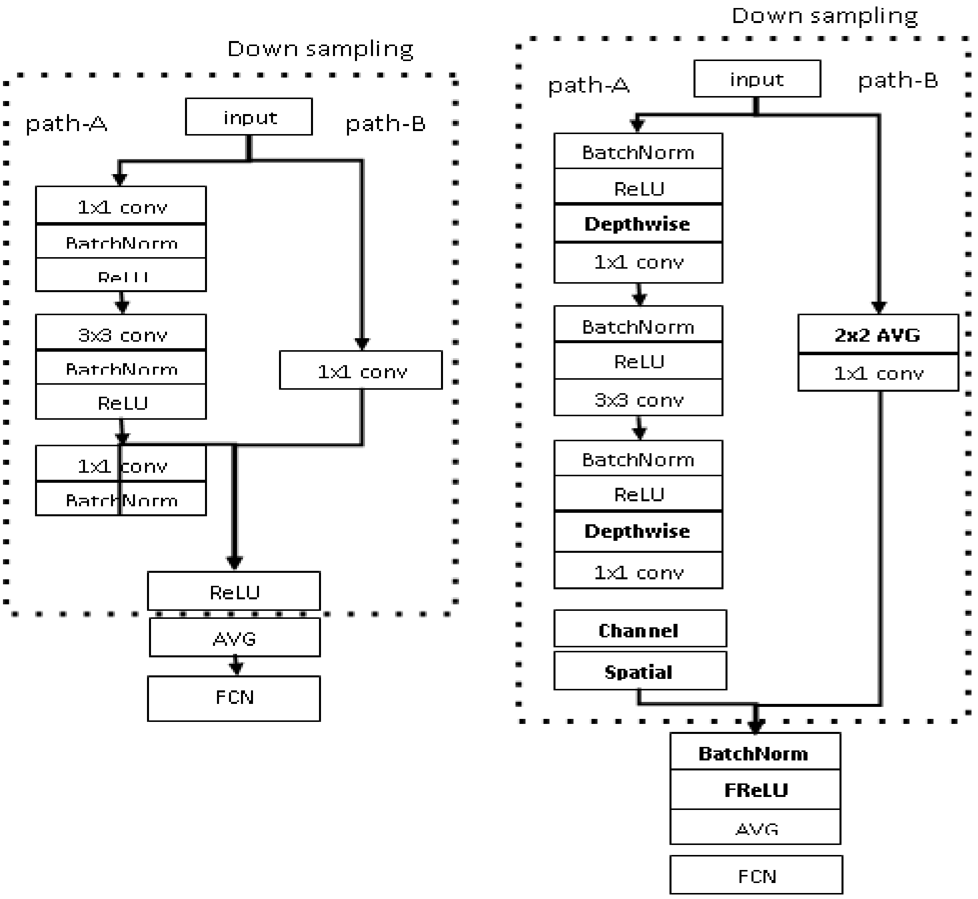
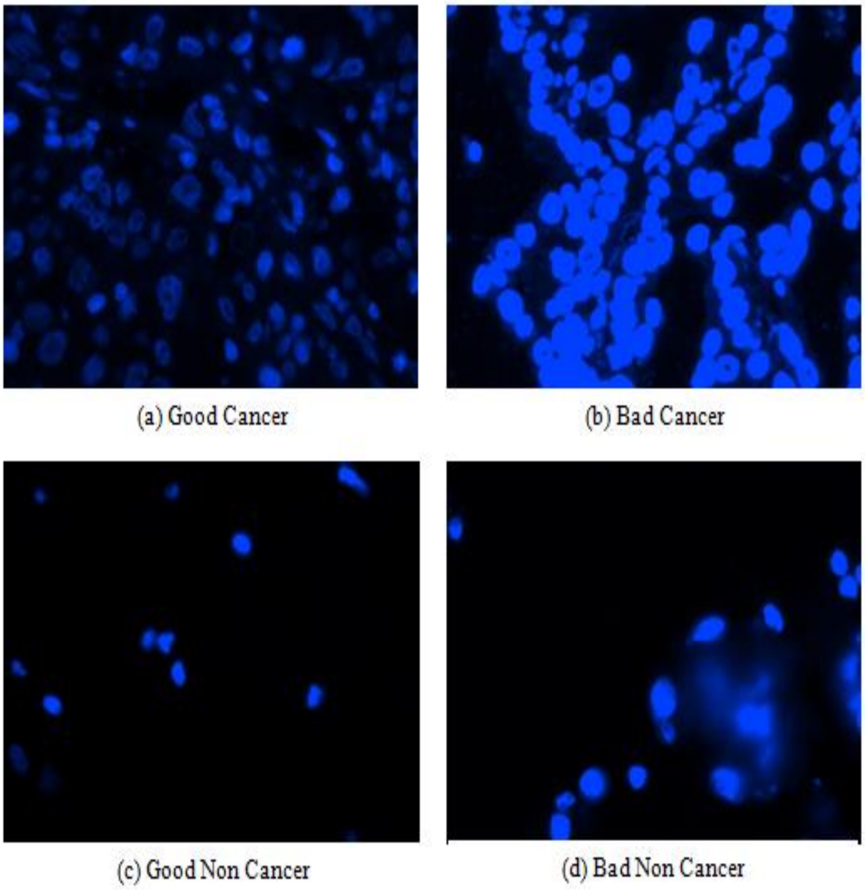


Figure 6. ResNet & proposed model

EXPERIMENT

Data Set

The data were received from the Catholic University of Korea Yeouido St. Mary's Hospital. The size of the data is 1,280x960, the data are DAPI images consisting of 3 channels, Train: 28,614, Validation: 3,577, Test: 3,578, and the total number of images is 35,769. As shown in Figure 7., the data set was classified into cancer and non-cancer data depending on the cells in the images which were cancer cells or normal cells, was divided into good images into cases where the cells in the images were clear or the shapes of cells were visible and bad images in cases where cells were not clear. However, in this study, good cancer and bad cancer were used as cancer, while good non-cancer and bad non-cancer were used as non-cancer, in order to determine the classification into cancer and non-cancer.


(a) Good Cancer

(b) Bad Cancer

(c) Good Non Cancer

(d) Bad Non Cancer

FIGURE 7: (a) Good Caner, (b) Bad Cancer, (c) Good Non Cancer, (d) Bad Non Cancer

TABLE 1: The identification rule of data labeling

Category		**Description**
Train	Cancer	Good Cancer , Bad Cancer
	Non Cancer	Good Non Cancer, Bad Non Cancer
Validation	Cancer	Good Cancer , Bad Cancer
	Non Cancer	Good Non Cancer , Bad Non Cancer
Test	Cancer	Good Cancer , Bad Cancer
	Non Cancer	Good Non Cancer , Bad Non Cancer

Learning Environment

In conducting this study, Windows 10 was used as the OS environment, and the model was trained and evaluated using NVIDIA GeForce RTX 3070 (8GB). The languages used for model development are Python 3.8.8 and Tensorflow 2.8.0, and finally, the CUDA version is 11.6.

TABLE 2: The Initialization parameters of training

Parameter	**Values**
Input	320 x 240
Batch Size	8
Optimizer	Adam
Learning rate	1e-4
Epochs	20
Early stopping	Patience 5

Comparison with Existing Models

ResNet, ResNetC, ResNetD, and ResNext were learned for performance comparison, and Accuracy (4), Precision (5), Recall (6), and F1 Score (7) were used as performance evaluation indexes. Accuracy is the percentage of cases where correct answers were given to all cases, but a problem with Accuracy is that it is determined by the ratio of correct answers no matter whether the correct answers are positive or negative so that proper results cannot be obtained when the ratio of negative answers is very high in the data. Therefore, Precision and Recall were used as indicators to evaluate whether answers are properly classified. Precision is the ratio of cases where the predicted values and the actual values of subjects who precited positively matched with each other and recall is the ratio of cases where the predicted values and actual values of actually positive facts match with each other as positive values. However, since Precision and Recall are indexes of opposite concepts, there is a limit to sufficiently expressing performance. Therefore, F1 Scores, which are the harmonic means of Precision and Recall, are used. The F1 Scores have values in a range of 0 to 1, and increase as Precision and Recall become similar to each other. The higher the F1 Score, the better the model's performance. Table 3. shows the performance evaluation between the proposed model and the comparative model. The Accuracy of the 'ResNet-ours' model proposed in this study was 82.83%, the Precision was 83%, the Recall was 84.14%, and the F1 Score was 83.56%, indicating that the performance of the proposed model is higher than that of the comparison model.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6)$$

$$\text{F1 Score} = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (7)$$

TABLE 3: The Performance of Proposed Model and Conventional Model

Model	**Accuracy**	**Precision**	**Recall**	**F1score**
ResNet(50)	0.8148	0.8005	0.8558	0.8272
ResNetC(50)	0.8119	0.8062	0.8409	0.8232
ResNetD(50)	0.8155	0.7930	0.8697	0.8296
ResNext(50)	0.7943	0.8060	0.7902	0.7980
ResNet-ours(50)	0.8283	0.8300	0.8414	0.8356

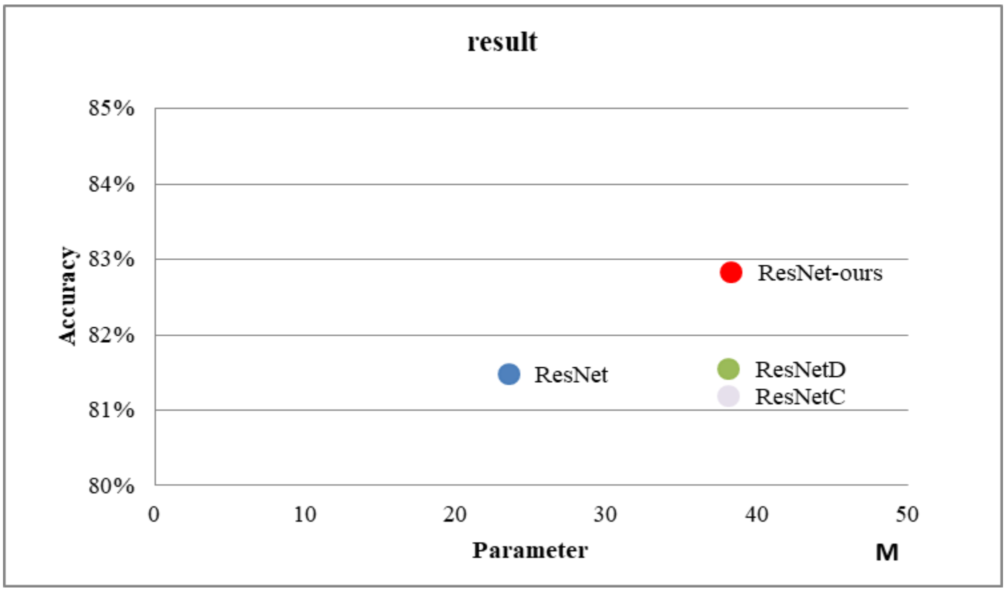


FIGURE 8: The Result Graph of Proposed Model and Conventional Model

CONCLUSIONS

In this study, to solve the problem of gradient loss, the Pre-activation structure made from the bottleneck structure of ResNet by carrying out batch normalization and changing the position of the activation function was applied, Depthwise Separable Convolution was applied through Depthwise Convolution, Self-attention Module was applied to find more characteristic Feature Maps, and finally, the FReLU function was applied. Thereafter, performance higher than that of existing ResNet-based models was shown through experiments. This model is expected to become a better model if only cancer cells are detected in the image through the classification results hereafter.

ACKNOWLEDGEMENTS

This work was supported by the Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, and Forestry(IPET) and Korea Smart Farm R&D Foundation(KoSFarm) through Smart Farm Innovation Technology Development Program, funded by the Ministry of Agriculture, Food and Rural Affairs(MAFRA) and Ministry of Science and ICT(MSIT), Rural Development Administration(RDA) (grant number: 421024-04). And this work was supported by the Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, and Forestry(IPET) funded by the Ministry of Agriculture, Food and Rural Affairs(MAFRA) (grant number: 321092-03-1-HD030).

REFERENCES

1. www.yoonsupchoi.com/2017/08/24/ai-medicine-6/, Aug 24 (2017)
2. Hong, J. Y., Park, S. H., & Jung, Y. J. (2020). Artificial intelligence based medical imaging: An overview. Journal of radiological science and technology, ISSN: 2288-3509(Print); 2384-1168(Online), Korean Society of Radiological Science, 43(3), 195-208. doi : https://doi.org/10.17946/jrst.2020.43.3.195
3. Yadav, S. S., & Jadhav, S. M. (2019). Deep convolutional neural network based medical image classification for disease diagnosis. Journal of Big Data, ISSN: 2196-1115, Springer, 6(1), 1-18. doi: https:// doi.org/ 10.1186/ s40537-019-0276-2
4. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, ISSN: 1063-6919, IEEE Computer Society, 770-778. doi : https://doi.org/ 10.1109/CVPR.2016.90W
5. Guo, Y., Li, Y., Wang, L., & Rosing, T. (2019, July). Depthwise convolution is all you need for learning multiple visual domains. In Proceedings of the AAAI Conference on Artificial Intelligence, ISSN : 2374-3468(Online); 2159-

5399(Print), AAAI-19, 33, 8368-8375. doi : https://doi.org/10.1609/aaai.v33i01.33018368

6. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, ISSN: 1049-5258, 30., 6000-6010. doi : https://doi.org/10.48550/arXiv.1706.03762
7. Wang, Q., Wu, B., Zhu, P.F., Li, P., Zuo, W., & Hu, Q. (2020). ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), ISSN: 1063-6919, IEEE Computer Society, 11531-11539. doi : https:// doi.org/ 10.1109/ CVPR42600.2020.01155
8. Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). Cbam: Convolutional block attention module. In Proceedings of the European conference on computer vision (ECCV), 3-19. doi : https://doi.org/ 10.1007/978-3-030-01234-2_1
9. Nair, V., & Hinton, G. E. (2010, January). Rectified linear units improve restricted boltzmann machines. In International Conference on Machine Learning(ICML)., 807-814
10. He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In Proceedings of the IEEE international conference on computer vision, ISSN: 1550-5499, Institute of Electrical and Electronics Engineers Inc, 1026-1034. doi : https://doi.org/ 10.1109/ ICCV.2015.123
11. Ma, N., Zhang, X., & Sun, J. (2020, August). Funnel activation for visual recognition. In European Conference on Computer Vision, 351-368. doi : https://doi.org/10.1007/978-3-030-58621-8_21
12. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition, ISSN: 1063-6919, IEEE Computer Society, 2818-2826. doi : https://doi.org/10.1109/CVPR.2016.308
13. He, T., Zhang, Z., Zhang, H., Zhang, Z., Xie, J., & Li, M. (2019). Bag of tricks for image classification with convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, ISSN: 1063-6919, IEEE Computer Society, 558-567. doi : https://doi.org/ 10.1109/ CVPR. 2019.00065
14. He, K., Zhang, X., Ren, S., & Sun, J. (2016, October). Identity mappings in deep residual networks. In European conference on computer vision, Springer, 630-645. doi : https://doi.org/ 10.1007/978-3-319-46493-0_38
15. Chollet, F. (2017). Xception: Deep Learning with Depthwise Separable Convolutions. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), ISSN: 1063-6919, IEEE Computer Society, 1800-1807. doi : https://doi.org/10.1109/CVPR.2017.195
16. Lee, S., Kim, Y. J. and Choi, Y. J. (2018). Does She Advance Her Development in The Face of Cancer? A Structural Equation Model of Posttraumatic Growth after Diagnosed with Cancer. International Journal of Advanced Nursing Education and Research, vol.3, no.2, Nov. 2018, pp.1-10, doi: https:// doi.org/ 10.21742/IJANER.2018.3.2.01
17. Sahamijoo, A., Piltan, F., Jaberi, S. M., Sulaiman, N. B. (2015). Prevent the Risk of Lung Cancer Progression Based on Fuel Ratio Optimization. International Journal of u - and e - Service, Science and Technology, NADIA, vol.8, no.2, Feb. (2015), pp.45-60, doi:10.14257/ijunnesst.2015.8.2.05.