비젼트랜스포머를 이용한 유방암 Xray 이미지 분류법

겔란 아야나*, 최세운*
금오공과대학교*

Vision transformer-based transfer learning for mammographic breast mass detection

Gelan Ayana*, Se-woon Choe*
Kumoh National Institute of Technology*

Abstract - Breast mass identification is the most important task doctors do during breast cancer early diagnosis using mammogram. However, it is challenging to identify at early stage whether a breast mass is benign or malignant. Hence, convolutional neural networks (CNNs) have been utilized to overcome the challenge and brought about practical improvements. Nevertheless, CNNs have drawbacks, such as focusing on a specific area of image overlooking the whole information on the mammogram image and also suffer from computational complexity due to the multiple convolution. Lately, vision transformers are introduced to the computer vision realm as a method that overcame these drawbacks of CNNs resulting in better or comparative performance for natural image classification. This success in natural image domain have not been explored carefully for medical image domain. Therefore, in this study, we propose the first vision transformer-based transfer learning for mammographic breast mass image classification.

1. Introduction

Breast cancer took over lung cancer being the first cause of death from cancer[1]. Early tumor detection using mammography has been reported to decrease the rate of death from breast cancer by 40% [1]. The radiologists look for the breast mass tissue to identify the tumor on mammogram [2]. However, it is a challenging task to determine whether the breast mass tissue from mammogram has a benign or malignant tumor. Thus, more than two experienced radiologists consensus is used in the conventional setting. The introduction of convolutional neural network (CNN) was taken as an opportunity to be utilized in improving early detection of breast cancer via mammography resulting in a few Food and Drug Administration (FDA) cleared artificial intelligence (AI) tools to assist radiologists decision making [2]. Nevertheless, CNNs are computationally complex warranting for high computing power as the size of data increase due to the multiple convolution tasks at different layers of the network [2]. Moreover, in case of mammogram image analysis, CNNs focus on a specific area (tumor suspected area) ignoring the other parts of the image, which lead CNNs to lose some important information that rather could be found by looking at the whole image at once Recently, vision transformers (ViTs) became popular in the computer vision field outperforming CNNs for natural image classification task [3]. ViTs outperformed the state-of-the-art CNN models with low computational complexity and solving the CNNs drawback of focusing on a specific area of image for classification. In this paper, we will explore the application of ViTs in classifying mammogram mass images for the early breast cancer detection.

2. Methods and results

2.1 The proposed method

The proposed method fine tune the ViTs variant called vitb_16, which is vision transformer base model with input patch size of 16x16 pixels, as proposed in Dosovitskiy et al, [3]. The original ViTs model output layer has been replaced with a flatten layer, batch normalization layer, and the final fully connected layer. The model was trained for 70 epochs using Adam optimizer with a starting learning rate of 0.0001.

2.1.1 Dataset

We obtained a mammogram breast mass images from the publicly available Mendeley mammogram breast dataset called Digital Database for Screening Mammography (DDSM) [4]. The dataset has a total of 13128 images categorized as benign and malignant with 7158 malignant and 5970 benign breast mass images. We have categorized the image into training, validation, test sets using 7:2:1 ratio for training, validation, and test, respectively.

Table 1. Dataset distribution.

DDSM	Benign	Malignant	Total
Training	4151	5064	9215
Validation	1239	1361	2600
Test	580	733	1313

2.2 Results

The proposed ViTs method provided high performance results in classifying the DDSM dataset with test accuracy of 100% and area under receiver operating curve (AUC) of 1. Furthermore, the proposed method provided a precision of 1, recall of 1, and F1-score of 1. The learning curve (Figure 1) and confusion matrix (Figure 2) to show how the training and testing of the proposed method are provided. The ViTs advantage of looking at the whole image at once in contrast to the CNNs that need many convolutions and layers to see only the tumor area has been shown in Figure 1, where ViTs performed well at early epochs.

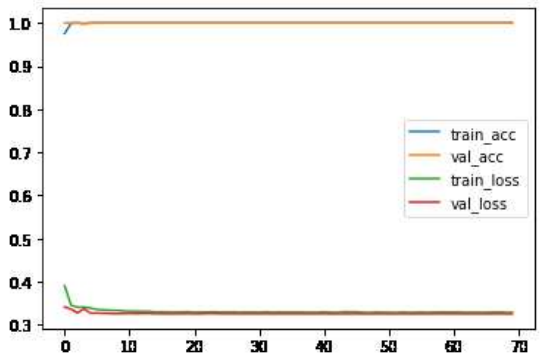

Figure 1. Learning curve of the proposed method.

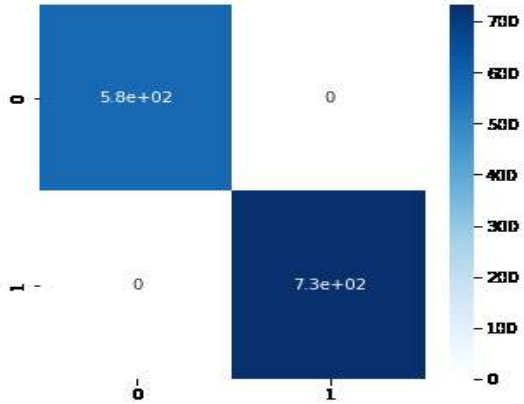

Figure 2. Confusion matrix of the proposed method.

2.2.1 Results discussion and future directions

The proposed ViTs model for classification of breast mammogram mass shown a high performance, classifying all the data accurately. ViTs performed well from the beginning at early epochs as can be seen in Figure 1. In the future, we will compare our result with state-of-the-art CNN models and also use different dataset to see the generalization power of ViTs.

3. Conclusion

In this paper, we have shown the first application of vision transformers for the classification of mammogram breast mass and recorded a superior performance This could be refined further and be of vital support in the early breast cancer diagnosis.

Acknowledgement

This research was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) [NRF-2019R1F1A1062397] and the Brain Korea 21 FOUR Project (Dept. of IT Convergence Engineering, Kumoh National Institute of Technology).

[References]

[1] H. Sung et al., "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," CA. Cancer J. Clin., vol. 71, no. 3, pp. 209-249, May 2021.

[2] G. Ayana, J. Park, and S. Choe, "Patchless Multi-Stage Transfer Learning for Improved Mammographic Breast Mass Classification," Cancers (Basel)., vol. 14, no. 5, p. 1280, Mar. 2022.

[3] A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," ICLR 2021, Oct. 2020.

[4] Heath, M.; Bowyer, K.; Kopans, D.; Kegelmeyer, P.; Moore, R.; Chang, K.; Munishkumaran, S. Current Status of the Digital Database for Screening Mammography. In Computational Imaging and Vision; Springer: Dordrecht, The Netherlands, 1998.