EfficientNetV2 기반 결핵 검출을 위한 전이 학습 활용

아흐마드 이자즈, 신석주*
컴퓨터공학과, 조선대학교

ahmadijaz@chosun.kr, *sjshin@chosun.ac.kr(교신저자)

Leveraging Transfer Learning in EfficientNetV2-based Tuberculosis Detection

Ijaz Ahmad, Seokjoo Shin
Dept. of Computer Engineering, Chosun University

Abstract

Training deep learning (DL) models is compute intensive task and requires massive datasets. Despite the improvements of DL models in natural images classification, their widespread adoption in healthcare sector is obstructed by the limited availability of medical data. Transfer leaning (TL) is one of the techniques that circumvents these problems. It provides rapid progress during training and improves model generalization on smaller dataset. In this study, we extend the applications of a recently proposed lightweight convolutional neural network (CNN) called EfficientNetV2 for tuberculosis (TB) detection in chest X-ray (CXR) images. Specifically, four variants (B0-B3) of the model were implemented and fine-tuned on publicly available Shenzhen CXR images dataset. The models performance was analyzed in terms of accuracy and sensitivity measures. The simulation analysis showed that without TL, EfficientNetV2-B1 achieved the best accuracy and sensitivity scores of 85%. These scores were then improved through the use of pre-trained models and reached to 89% accuracy and 94% sensitivity scores for EfficientNetV2-B3.

Ⅰ. Introduction

According to the World Health Organization (WHO) report [1], Tuberculosis (TB) remains one of the leading cause of death from a single infectious agent. The disease spread and mortality is preventable when early detected [2]. For TB screening, the WHO recommends CXR images due to its relative low cost and wide availability. With the recent success of ML in the field of computer vision, automatic aided diagnosis (CAD) systems have emerged to assist doctors and practitioners. However, this progress is hindered by the limited available medical data. Transfer learning is a machine learning (ML) technique where what has been learned on a larger dataset is repurposed to improve generalization in another related smaller dataset. It is a popular technique in certain domains such as automatic medical image analysis where it is hard and expensive to get a dataset of sufficient size to train a network from scratch with random initialization [3].

Several DL models have been proposed for diagnosis of TB. Among them, EfficientNet models provide parameter efficiency with better training and inference speed. For example, EfficientNetV1 [4] based model has achieved 89.92% accuracy in [5]. Their accuracy has been improved to 94% with TL in [2]. For the most recent EfficientNetV2 [6], accuracy of 89.52% has been achieved without TL and pre-processing in [7].

In this study, we proposed a deep learning model based on EfficientNetV2 for TB screening in CXR images. For this purpose, we have analyzed EfficientNetV2 B0 – B3 models. The motivation behind our model choice is that EfficientNetV2 provides faster training time, and better parameters efficiency. In order to improve our model performance, we have used pre-trained EfficientNetV2 models to benefit from TL.

Ⅱ. Methods

EfficientNet (EfficientNetV1 [4]) is a family of lightweight convolutional models that are optimized for number of parameters and floating-point operations (FLOPs). Their superiority over existing CNN models can be attributed to the use of depthwise convolutions. However, such convolutions often cannot fully utilize modern accelerators; therefore, they have a main limitation in terms of training or inferencing speed. The recently proposed EfficientNetV2 [6] improves their training speed while maintaining the parameter efficiency. Specifically, EfficientNetV2 provides three solutions. First, for better training speed, it proposed to adjust the image size and regularization progressively during training. Second, EfficientNetV2 proposed a non-uniform scaling strategy to add more layers to later stages gradually as opposed to EfficientNetV1 that equally scales up all stages by using a simple compound scaling rule. Finally, to fully utilize modern

TABLE I. Performance analysis of the EfficientNetV2 variants with and without transfer learning for TB detection in Shenzhen dataset. The measures values are reported as the mean (black color) and standard deviation (gray color) for 3 runs. (Best performance is in bold for each case).

Metrics	Without Transfer Learning				With Transfer Learning			
	B0	B1	B2	B3	B0	B1	B2	B3
Accuracy	83 ± 0.014	**85 ± 0.005**	83 ± 0.008	84 ± 0.009	89 ± 0.008	88 ± 0.014	89 ± 0.014	**89 ± 0.005**
Sensitivity	83 ± 0.014	**85 ± 0.021**	81 ± 0.031	83 ± 0.028	91 ± 0.024	87 ± 0.021	93 ± 0.019	**94 ± 0.029**

accelerators, EfficientNetV2 proposed that FusedMBConv in early stage can improve training speed with a small overhead on parameters and FLOPs.

Ⅲ. Results

Dataset: In this study, we have used a publicly available CXR images dataset called Shenzhen (SH) China dataset [8]. It consists of 662 CXR images in total. In simulations, we have used only 326 images from each class for balance distribution. The dataset split was 80% for training, 10% for validation and 10% for testing. In addition, the input images were pre-processed as: (1) all black borders and regions on the edges of images were cropped and (2) the images were resized from the center to meet the models input size requirements.

Metrics: In analysis, we considered samples with TB as positive and healthy samples as negative classes. The true positives (TP) is the number of observations belonging to the positive class and are correctly classified while false negatives (FN) are the misclassified ones. Similarly, true negatives (TN) is the number of observations belonging to the negative class and are correctly classified as such and misclassified as positive class are false positives (FP). For the models performance evaluation, we have considered two measures namely, accuracy, and sensitivity. The performance metrics are defined as:

$$\begin{cases} \text{Accuracy} = (\text{TP} + \text{TN})/(\text{TP} + \text{TN} + \text{FP} + \text{FN}) \\ \text{Sensitivity} = (\text{TP})/(\text{TP} + \text{TN}) \end{cases} \quad (1)$$

Accuracy measures the total number of correct predictions (TP + TN) made by the model, which is the ratio of correct predictions to total predictions. It is an important metric when FN and FP have similar costs. However, for disease diagnosis, the occurrence of FN is intolerable and identifying the positives is crucial [7]. For this purpose, sensitivity is the metric that measures performance of a model in terms of how fewer number of FN are predicted.

Training Setup: When the models were trained from scratch number of epochs was set to 150 and the initial learning rate (lr) was set to 0.1. For regularization, the lr was reduced by a factor of 10 when validation accuracy stopped improving for 20 epochs. In addition, we have used early stopping criteria once the model validation accuracy stopped improving for 60 epochs. For transfer learning, we have trained the models in two steps. First, the ConvNet layers of a model were frozen and only classifier was trained for 40 epochs. In the second step, the whole model was fine-tuned for 40 epochs. In both steps, the same lr=10e-5 was used. Throughout our simulations, Stochastic Gradient Descent (SGD) and batch size=32 were used. The training dataset was augmented using random flip, rotation, zoom and translation.

Table 1. presents performance of EfficientNetV2 variants in terms of accuracy, and sensitivity for TB detection in X-ray images. When the model was trained from scratch, then the best accuracy and sensitivity scores of 85% were achieved (B1). On the other hand, when using pre-trained model weights the accuracy and sensitivity scores improved to 89% and 94%, respectively (B3). Overall, for each variant, the metrics scores improved with the transfer learning.

Ⅳ. Conclusion

This study extended the applications of EfficientNetV2 for TB detection in CXR images. In the experiments, TL was utilized to improve the performance of the classification model. Particularly, the analysis showed that with TL, the larger models achieved better performance than smaller ones.

ACKNOWLEDGMENT

This research is supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2018R1D1A1B07048338).

References

[1] “Global tuberculosis report 2021.” https://bit.ly/3AQASEu (accessed Aug. 28, 2022).

[2] M. Oloko-Oba and S. Viriri, “Ensemble of EfficientNets for the Diagnosis of Tuberculosis,” *Computational Intelligence and Neuroscience*, vol. 2021, pp. 1–12, Dec. 2021, doi: 10.1155/2021/9790894.

[3] I. Ahmad and S. Shin, “An Approach to Run Pre-Trained Deep Learning Models on Grayscale Images,” in *Inter. Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, Jeju, Korea, Apr. 2021, pp. 1–4.

[4] M. Tan and Q. V. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,” Sep. 2020, [Online]. Available: http://arxiv.org/abs/1905.11946

[5] K. Munadi, K. Muchtar, N. Maulina, and B. Pradhan, “Image Enhancement for Tuberculosis Detection Using Deep Learning,” *IEEE Access*, vol. 8, pp. 217897–217907, 2020, doi: 10.1109/ACCESS.2020.3041867.

[6] M. Tan and Q. V. Le, “EfficientNetV2: Smaller Models and Faster Training,” Jun. 2021, [Online]. Available: http://arxiv.org/abs/2104.00298

[7] I. Ahmad and S. Shin, “A Perceptual Encryption-Based Image Communication System for Deep Learning-Based Tuberculosis Diagnosis Using Healthcare Cloud Services,” *Electronics*, vol. 11, no. 16, p. 2514, Aug. 2022, doi: 10.3390/electronics11162514.

[8] S. Jaeger, S. Candemir, S. Antani, Y.-X. J. Wáng, P.-X. Lu, and G. Thoma, “Two public chest X-ray datasets for computer-aided screening of pulmonary diseases,” *Quantitative Imaging in Medicine & Surgery*, vol. 4, no. 6, 2014.