Evaluation of Classification Performance of Inception V3 Algorithm for Chest X-ray Images of Patients with Cardiomegaly

Woo-Yeon Jeong[1], Jung-Hun Kim[2], Ji-Eun Park[3], Min-Jeong Kim[1], Jong-Min Lee[4,*]

[1]Department of Biomedical Engineering, Kyungpook National University
[2]Bio-Medical Research institute, Kyungpook National University Hospital
[3]Nonlinear Dynamics Research Center, Kyungpook National University
[4]Department of Radiology, School of Medicine, Kyungpook National University

Received: May 28, 2021. Revised: August 20, 2021. Accepted: August 31, 2021.

ABSTRACT

Cardiomegaly is one of the most common diseases seen on chest X-rays, but if it is not detected early, it can cause serious complications. In view of this, in recent years, many researches on image analysis in which deep learning algorithms using artificial intelligence are applied to medical care have been conducted with the development of various science and technology fields. In this paper, we would like to evaluate whether the Inception V3 deep learning model is a useful model for the classification of Cardiomegaly using chest X-ray images. For the images used, a total of 1026 chest X-ray images of patients diagnosed with normal heart and those diagnosed with Cardiomegaly in Kyungpook National University Hospital were used. As a result of the experiment, the classification accuracy and loss of the Inception V3 deep learning model according to the presence or absence of Cardiomegaly were 96.0% and 0.22%, respectively. From the research results, it was found that the Inception V3 deep learning model is an excellent deep learning model for feature extraction and classification of chest image data. The Inception V3 deep learning model is considered to be a useful deep learning model for classification of chest diseases, and if such excellent research results are obtained by conducting research using a little more variety of medical image data, I think it will be great help for doctor's diagnosis in future.

Keywords: Cardiomegaly, Artificial Intelligence, Deep learning, Chest X-ray Image, Classification

I. INTRODUCTION

심장비대증(Cardiomegaly)은 심장 내부 공간의 확장이나 심실비대에 의해 심장이 커진 상태를 의미하며 초음파나 흉부 X선 영상의 판독으로 진단할 수 있다[1]. 그러나 의료 영상판독 시 전문의의 숙련된 기술이나 영상의 질에 따라 진단 결과가 달라질 수도 있으며[2,3], 조기에 발견되지 못하면 심각한 합병증을 초래할 수 있다[4,5].

최근에는 인공지능을 이용한 딥러닝 알고리즘을 의료에 접목시키는 연구들이 많이 진행되고 있다[6]. 그 중에는 영상의학 분야에 딥러닝 알고리즘을 이용한 이미지 분류 연구들도 많이 진행되고 있다[7,8]. 현재 이미지 분류를 이용한 딥러닝 알고리즘에는 VGGNet, ResNet, GoogleNet 등 우수한 분류 성능을 보여준 여러 알고리즘이 있고 지속적으로 새로운 알고리즘이 개발되어지고 있다[9]. 이렇게 다양한 딥러닝 알고리즘을 이용한 이미지 분류 대회에서 가장 우수한 분류 성능을 보인 GoogleNet(Inception) 알고리즘은 구조를 점차 개선하여 Inception V3 모델을 개발하였고 의료 영상을 이용한 이미지 분류에서

* Corresponding Author: Jong-Min Lee E-mail: jonglee@knu.ac.kr Tel: +82-53-950-7956
Address: Department of Radiology, School of Medicine, Kyungpook National University, Dongin-dong, Jung-gu, Daegu, Korea

도 우수한 결과를 보여주었다[10,11]. 딥러닝 알고리즘에 의료 영상을 이용한 분류 연구에서는 흉부 X선 영상이 많이 사용되어지고 있었으며 특히 폐에 대한 분류 연구들이 대부분이었다[12,13]. 그 중 Inception V3 알고리즘을 이용하여 온라인에서 제공되어지는 흉부 X선 영상의 폐 질환에 대한 분류 연구에서는 우수한 결과를 보였다[10].

현재 딥러닝 알고리즘을 이용하여 흉부 X선 영상에서 폐 질환에 대한 분류 연구는 많이 있으나 심장 질환에 대한 영상 분류 연구는 많이 없다. 본 연구에서는 Inception V3 알고리즘을 이용하여 직접 수집한 흉부 X선 이미지 분류를 진행하였을 때 심장비대증 의료 영상 분류에 적합한 알고리즘인지 알아보고자 한다.

II. MATERIAL AND METHODS

1. 실험 재료

1.1 프로그래밍 언어 & 프레임 워크

Table 1은 연구에 사용된 프로그래밍 언어와 프레임 워크를 나타내며 딥러닝 모델링을 구현하기 위해 파이썬(Python, version 3.8.5) 언어를 사용하였다. 딥러닝 모델링 구현을 위해 프레임 워크로는 케라스(Keras, version 2.4.3)를 적용하였다.

Table 1. Programming Language and Framework

Components	Version
Python	3.8.5
Keras	2.4.3

1.2 가상환경 & 구현환경

Table 2는 연구에 사용된 가상환경과 구현환경을 나타내며 가상환경으로는 아나콘다(Anaconda, version 2020.11, Continuum Analytics, Texas, USA)를 사용하였으며, 구현환경으로는 주피터 노트북(Jupyter notebook, version 6.1.4)을 이용하였다.

1.3 컴퓨터 사양

Table 3은 연구에 사용된 컴퓨터 사양을 나타내며 CPU : Intel (R) Core (TM) i7-8700k CPU @3.70GHZ, RAM : 32.0GB, GPU : NVIDIA GeForce GTX 1080 Ti를 사용하였다.

Table 2. Virtual Environment and Implementation Environment

Components	Version
Anaconda	2020.11
Jupyternotebook	6.1.4
Tensorflow	2.4.1
Numpy	1.19.2
Matplotlib	3.3.2

Table 3. Computer specifications

Components	Specification
CPU	Intel (R) Core (TM) i7-8700k
Speed	3.70 GHz
RAM	32.0GB
GPU	NVIDIA Geforce GTX1080 Ti
OS	Windows 10

1.4 Inception V3 딥러닝 모델

Inception V3는 48개의 계층으로 구성된 컨볼루션 신경망이며 여러 가지 엄격한 제약 조건에서도 잘 수행될 수 있는 Fig. 1의 구조로 설계되어 있다[14,15].

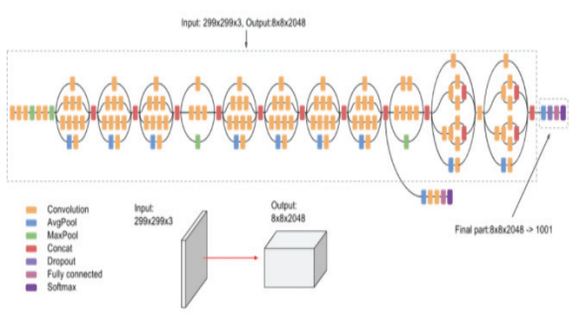

Fig. 1. The structure of Inception V3.

Inception V3는 노드 간의 연결을 줄이고, 행렬연산은 Dense 연산을 하는 구조로 이미지 특징맵을 추출하기 위하여 Convolution 층을 1x1, 3x3, 5x5로 구성하여 연결하였고 Max pooling을 이용하여 연산을 수행한다. 이때 수행하는 연산량이 많아서 1x1

Convolution을 부분마다 넣어 차원을 감소시킨 후 3x3과 5x5 Convolution 연산을 수행하게 되므로 연산의 양을 축소되며 효율적인 연산이 가능하다. 이러한 구조로 Inception V3는 연산의 복잡도를 줄이고 파라미터 수를 적게 사용함으로써 학습 속도를 향상시킨다는 장점이 있다[16,17].

1.5 데이터 세트(Data set)

데이터 세트는 경북대학교병원의 임상시험심사위원회(IRB; Institutional Review Board, 2020-08-017)의 승인 후 병원 내 20세 이상에서 95세 미만의 환자들을 대상으로 2010년 1월부터 2020년 12월까지 정상 진단을 받은 환자의 흉부 X선 영상 526장과 심장비대증 진단을 받은 환자의 흉부 X선 영상 500장을 사용하였다. 데이터 세트는 총 1026장의 흉부 X선 영상 TIFF 이미지 파일로 구성되어 있다.

2. 실험 방법

직접 수집한 흉부 X선 영상을 이용한 심장비대증의 분류에 대한 성능평가를 위하여 Inception V3 모델을 사용하였다. Inception V3 딥러닝 모델링 구현 시 흉부 X선 영상 파일을 데이터 세트로 사용하였다.

2.1 데이터 세트의 분류

직접 수집한 흉부 X선 영상 데이터 세트를 사용하였다. 우선 전체 1026개의 흉부 X선 영상 데이터 중 정상 심장 영상은 526개로 라벨링을 1, 심장비대증 영상은 500개로 라벨링을 0으로 작업을 하였다. 라벨링을 완료한 흉부 X선 영상은 Table 4와 같이 학습 데이터 세트, 검증 데이터 세트, 테스트 데이터 세트로 분류하여 사용하였다. 정상 심장의 경우 흉부 X선 영상 526개 중 학습 데이터 세트는 226개로 딥러닝 모델을 학습시키기 위해서 사용되며, 검증 데이터 세트는 200개로 학습시킨 딥러닝 모델이 제대로 학습되었는지 확인하기 위해 사용되어지며, 테스트 데이터 세트는 100개로 최종적으로 모델을 평가하기 위해 사용된다. 심장비대증의 경우 흉부 X선 영상 500개 중 학습 데이터 세트 250개, 검증 데이터 세트 150개, 테스트 데이터 세트 100개로 분류하였다.

Table 4. Classification of Data set

Components	Normal	Cardiomegaly
Training Data set	226	250
Validation Data set	200	150
Test Data set	100	100

2.2 딥러닝 모델링 (Deep learning modeling)

딥러닝 모델링은 Inception V3 모델의 기본 구조를 사용하였으며, 활성화 함수는 Sigmoid를 적용하였다. 최적화 알고리즘은 RMSProp를 적용하였으며 배치 사이즈는 32로 설정하였고, 학습 에포크(Epoch)는 40으로 설정하여 흉부 X선 영상의 정상 심장과 심장비대증 분류에 대한 딥러닝 모델링을 구현하였다. 여기서, 배치 사이즈는 한 번의 반복(Iteration)에 이용되는 이미지의 수를 얼마만큼 나눠서 넣을지 지정하는 값을 의미하고, 학습 에포크는 전체 학습 셋이 신경망을 통과한 횟수를 의미한다. 연구에서 사용된 학습 에포크는 40이 넘어갈수록 오버피팅(Over-fitting)현상을 보여 에포크를 40으로 설정하였다. 활성화 함수는 입력받은 데이터를 다음 레이어에 어떤 방식으로 전달을 해주는 가를 의미하며 활성화 함수로 쓰인 Sigmoid는 입력데이터의 값에 따라 0~1의 값을 출력하는 S자형 함수를 의미한다. 본 연구에서는 Relu, Softmax, Tanh 함수를 적용하였을 때 보다 Sigmoid 함수를 적용하였을 때 가장 우수한 성능을 보여 Sigmoid 함수를 적용하였다. 최적화 알고리즘은 손실 함수를 최소화하는 방향으로 가중치를 갱신하는 함수를 의미하며 최적화 알고리즘으로 사용된 RMSProp는 학습하는 속도를 적절하게 수정하며 최적화하는 방식을 의미한다.

3. 평가 방법

Inception V3 모델을 이용하여 딥러닝 모델링의 학습 에포크마다 출력되는 결과 딥러닝 모델의 성능평가에 사용하였다. 평가하는 항목으로는 학습 모델링과 검증 모델링의 정확도(Accuracy), 손실도(Loss) 그리고 모델의 정상 심장과 심장비대증에 대한 분류 정밀도(Precision), 재현율(Recall), F1 스코어(F1 score)를 평가하였다.

3.1 딥러닝 모델의 성능평가

딥러닝 모델의 성능평가를 위하여 학습 모델링과 검증 모델링의 정확도와 손실도, 정밀도, 재현율, F1 스코어 값을 사용하였다. 여기서 정확도는 딥러닝 모델의 검증 모델링의 결과값인 예측값과 실측값을 대조할 때 정확한 정도를 의미하며 전체 데이터 수 중에 예측 결과값과 실제 실측값의 동일한 건수가 차지하는 비율이다. 손실도는 실측값과 예측값을 대조할 때 발생하는 오차의 정도를 의미하며 전체 데이터 수 중에 예측 결과값과 실제 실측값의 일치하지 않는 건수가 차지하는 비율이다. 정밀도는 예측을 심장비대증으로 한 예측값 중에 실제값이 심장비대증으로 일치한 비율이며 재현율은 전체 데이터 중에 심장비대증에 대해 찾아내는 비율 혹은 심장비대증이 아닌 값을 찾아내는 비율이다. F1 스코어는 정밀도와 재현율의 조화에 대한 평균비율을 의미한다.

III. RESULT

1. Inception V3 딥러닝 모델의 성능 평가

1.1 정확도 평가

Inception V3 딥러닝 모델을 이용한 흉부 X선 영상의 정상 심장과 비정상 심장 분류에 대한 정확도 평가를 진행하였다. Fig. 2는 학습 모델링과 검증모델링의 심장비대증의 유무에 따른 분류 정확도 평가한 그래프이다. 학습 모델의 정확도는 학습 초기인 에포크 1에서는 62.6%의 정확도를 나타내었지만 최종 에포크 40에서 96.0%의 정확도를 나타내었다. 검증 모델의 정확도의 경우 에포크 1에서 57.1%를 나타내었으나 최종 에포크 40에서 79.4%의 정확도를 보였다.

1.2 손실도 평가

Inception V3 딥러닝 모델을 이용한 흉부 X선 영상의 정상 심장과 비정상 심장 분류에 대한 손실도 평가를 진행하였다. Fig. 3는 학습 모델링과 검증모델링의 심장비대증의 유무에 따른 분류 손실도를 평가한 그래프이다. 학습 모델의 손실도의 경우 학습 초기인 에포크 1에서는 2.06%를 나타내었으나 최종 에포크 40에서 0.14%의 손실도를 보였다. 검증 모델의 손실도의 경우 에포크 1에서는 3.26%를 나타내었으나 최종 에포크 40에서 0.88%의 손실도를 보였다.

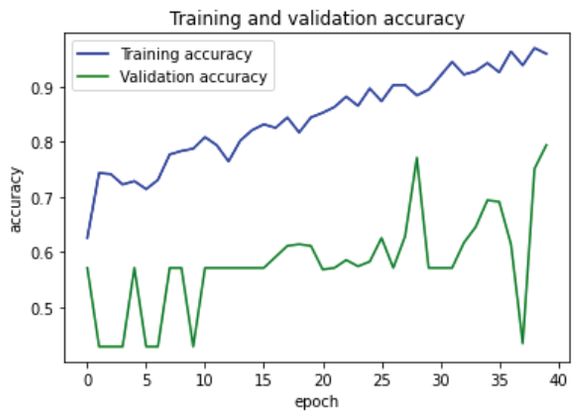

Fig. 2. Training and Validation Accuracy.

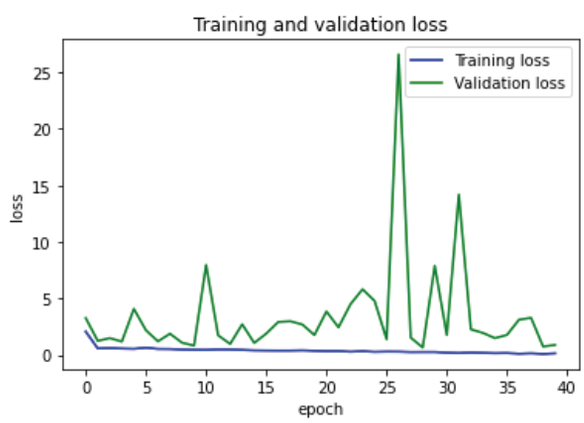

Fig. 3. Training and Validation Loss.

1.3 정밀도, 재현율, F1 스코어 평가

Inception V3 딥러닝 모델을 이용한 흉부 X선 영상의 정상 심장과 비정상 심장 분류에 대한 정밀도, 재현율, F1 스코어 평가를 진행하였으며 Table 5는 모델의 분류를 평가한 표이다. 정상 심장에 대한 분류 정밀도는 100%와 재현율이 87%로 F1 스코어가 93%인 결과를 보였고 비정상 심장인 심장비대증에 대한 분류 정밀도는 89%와 재현율이 100%로 F1 스코어가 94%인 결과를 보였다.

Table 5. Evaluation of Precision, Recall, F1 Score

Metric	Normal	Cardiomegaly
Precision	100(%)	89(%)
Recall	87(%)	100(%)
F1 Score	93(%)	94(%)

IV. CONCLUSION

본 논문에서는 인공지능 딥러닝 알고리즘인 Inception V3를 이용하여 직접 수집한 심장비대증 흉부 X선 영상 데이터가 사용된 딥러닝 모델링을 구현하였다. 실험을 통해 딥러닝 모델링을 적용한 정상 심장과 심장비대증의 분류에 있어 정확도와 손실도를 이용한 성능평가를 진행했다. 실험결과 Inception V3 딥러닝 모델은 성능평가에서 정상 심장과 심장비대증 분류에 우수한 성능을 나타내었다. 다만, 손실도 평가에서 검증 모델링의 손실도가 크게 증가하는 현상이 있었는데 이는 추후 연구 시 데이터의 수를 증대시키거나 모델의 구조를 더 간단하게 바꾼다면 해결이 가능한 부분이라고 생각되며 해당 연구에서 구현한 딥러닝 모델인 Inception V3는 흉부 X선 데이터를 이용한 분류에 있어 유용한 딥러닝 모델이 될 것이라고 사료된다. 흉부 X선 영상에 국한되지 않고 신체의 다른 부분 영상이나 CT 또는 MRI 영상처럼 다양한 의료 영상 데이터를 이용한 연구를 진행하여 이와 같은 우수한 연구결과를 얻게 된다면 향후 임상의의 진단 시 많은 도움을 줄 수 있을 것으로 기대된다.

Reference

[1] E. F. Philbin, R. Garg, K. Danisa, D. M. Denny, G. Gosselin, C. Hassapoyannes, "The Relationship Between Cardiothoracic Ratio and Left Ventricular Ejection Fraction in Congestive Heart Failure", Archives of Internal Medicine, Vol. 158, No. 5, pp. 501-506, 1998. http://dx.doi.org/10.1001/archinte.158.5.501

[2] J. C. Anderson, H. A. Baltaxe, G. L. Wolf, "Inability to show clot: one limitation of ultrasonography of the abdominal aorta", Radiology, Vol. 132, No. 3, pp. 693-696, 1979. http://dx.doi.org/10.1148/132.3.693

[3] H. J. Moon, E. K. Kim, J. S. Park, J. Y. Kwak, "Thyroid Ultrasound: Change of Inter-observer Variability and Diagnostic Performance after Training", Journal of Korean Society of Ultrasound in Medicine, Vol. 30, No. 1, pp. 23-28, 2011.

[4] E. D. Frohlich, "Left ventricular hypertrophy as a risk factor", Cardiology Clinics, Vol. 4, No. 1, pp. 137-144, 1986.

[5] D. Levy, K. M. Anderson, D. D. Savage, W. B. Kannel, J. C. Christiansen, W. P. Castelli, "Echocardiographically detected left ventricular hypertrophy: prevalence and risk factors: the Framingham Heart Study", Annals of Internal Medicine, Vol. 108, No. 1, pp. 7-13, 1988. https://doi.org/10.7326/0003-4819-108-1-7

[6] S. Lim, M. Lee, "A Computer Aided Diagnosis Algorithm for Classification of Malignant Melanoma Based on Deep Learning", Journal of the Korea Society of Digital Industry and Information Management, Vol. 14, No. 4, pp. 69-77, 2018.

[7] Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning", arXiv:1711.05525, 2017. https://arxiv.org/abs/1711.05225v3

[8] L. Yao, E. Poblenz, D. Dagunts, B. Covington, D. Bernard, K. Lyman, "Learning to diagnose from scratch by exploiting dependencies among labels", arXiv preprint arXiv:1710.10501. 2017. https://arxiv.org/abs/1710.10501

[9] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, R. M. Summers, "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases", In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2097-2106, 2017. https://doi.org/10.1109/CVPR.2017.369

[10] J. Y. Kim, S. Y. Ye, "Diagnostic Classification of Chest X-ray Pneumonia using Inception V3 Modeling", Journal of the Korean Society of Radiology, Vol. 14, No. 6, pp. 773-780, 2020. https://doi.org/10.7742/jksr.2020.14.6.773

[11] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, A. Rabinovich, "Going deeper with convolutions", In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1-9, 2015. https://arxiv.org/abs/1409.4842

[12] M. S. Ko, B. C. Jeong, D. G. Kim, C. Han, "Deep Learning Under Privileged Information for Pneumonia Detection", The Institute of Electronics and Information Engineers, Vol. 58, No. 3, pp. 67-73, 2021. https://arxiv.org/abs/1805.11614

[13] H. J. Song, E. B. Lee, H. J. Jo, S. Y. Park, S. Y. Kim, H. J. Kim, J. W. Hong, "Evaluation of Classification and Accuracy in Chest X-ray Images using Deep Learning with Convolution Neural Network", Journal of the Korean Society of Radiology, Vol. 14, No. 1, pp. 39-44, 2020. https://doi.org/10.7742/jksr.2019.14.1.39

[14] Google. Advanced Guide to Inception v3 on Cloud TPU. https://cloud.google.com/tpu/docs/inception-v3-advanced?hl=en

[15] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, "Rethinking the Inception Architecture for Computer Vision", arXiv preprint arXiv:1512.00567. 2015.

[16] J. K. Lee, S. J. Kim, N. J. Kwak, D. W. Kim, J. H. Ahn, "A Deep Learning Model for Judging Presence or Absence of Lesions in the Chest X-ray Images", The Journal of the Korean Institute of Information and Communication Engineering, Vol. 24, No. 2, pp. 212-218, 2020. https://doi.org/10.6109/jkiice.2020.24.2.212

[17] K. D. Song, M. C. Kim, S. H. Do, "The Latest Trends in the Use of Deep Learning in Radiology Illustrated Through the Stages of Deep Learning Algorithm Development", Korean Journal of Radiology, Vol. 80, No. 2, pp. 202-212, 2019. http://dx.doi.org/10.3348/jksr.2019.80.2.202

심장비대증 환자의 흉부 X선 영상에 대한 Inception V3 알고리즘의 분류 성능평가

정우연[1], 김정훈[2], 박지은[3], 김민정[1], 이종민[4,*]

[1]경북대학교대학원 의용생체공학과
[2]경북대학교병원 생명 의학 연구원
[3]경북대학교 비선형 동역학 연구소
[4]경북대학교 의과대학 영상의학교실

요 약

심장비대증은 흉부 X선 영상에서 흔히 보이는 질병 중 하나이지만 조기에 발견을 하지 못하면 심각한 합병증을 유발할 수도 있다. 이러한 점을 고려하여 최근에는 여러 과학기술 분야의 발전으로 인공지능을 이용한 딥러닝 알고리즘을 의료에 접목시키는 영상 분석 연구들이 많이 진행되고 있다. 본 논문에서는 Inception V3 딥러닝 모델을 흉부 X선 영상을 이용하여 심장비대증의 분류에 유용한 모델인지 평가하고자 한다. 사용된 영상의 경우 총 1026장의 경북대학교병원 내 정상 심장 진단을 받은 환자와 심장비대증 진단을 받은 환자의 흉부 X선 영상을 사용하였다. 실험결과 Inception V3 딥러닝 모델의 심장비대증 유무에 따른 분류 정확도와 손실도 결과값은 각각 96.0%, 0.22%의 결과값을 나타내었다. 연구결과를 통해 Inception V3 딥러닝 모델은 흉부 영상 데이터의 특징 추출 및 분류에 있어 우수한 딥러닝 모델인 것을 알 수 있었다. Inception V3 딥러닝 모델의 경우 흉부 질환의 분류에 있어 유용한 딥러닝 모델이 될 것으로 판단되며 조금 더 다양한 의료 영상 데이터를 이용한 연구를 진행하여 이와 같은 우수한 연구결과를 얻게 된다면 향후 임상의의 진단 시 많은 도움을 줄 수 있을 것으로 사료된다.

중심단어: 심장비대증, 인공지능, 딥 러닝, 흉부 X선 촬영영상, 분류

연구자 정보 이력

	성명	소속	직위
(제1저자)	정우연	경북대학교대학원 의용생체공학과	석사과정
(공동저자)	김정훈	경북대학교병원 생명 의학 연구원	박사
	박지은	경북대학교 비선형 동역학 연구소	박사
	김민정	경북대학교대학원 의용생체공학과	석사과정
(교신저자)	이종민	경북대학교 의과대학 영상의학교실	교수