

폐질환 의심 응급환자의 진단 과정 단축을 위한 AI기반 X-ray진단 시스템

정행복*, 최세연*, 김현수*

AI-based X-ray diagnostic system implementation to shorten the diagnosis process of emergency patients suspected of lung disease

Haengbok Chung*, Saeyoun Choi*, and Hieonn Kim*

요 약

응급실 방문 환자의 반 이상(56%)이 x-ray 촬영을 진행하고 있다. 하지만 응급실에서 흉부 x-ray 진단에 소요되는 시간은 약 1시간으로 골든타임 내에 적절한 응급 치료를 하는 것이 중요한 응급환자에게는 긴 시간이다. 이러한 시간 지연 문제를 해결하기 위해 12초 안에 x-ray사진만으로 질병을 진단하고 감염부위를 특징하는 시스템을 제안한다. 이를 구현하기 위해 Classification, Detection 기능 기반 Web Service를 제작하였고 Classification 성능을 향상시키기 위해 PGGAN을 이용하여 Synthetic Medical X-ray Data를 생성하여 데이터 셋의 Class Imbalance를 완화하여 진단 성능을 향상시키고 진단 시간을 단축하여, 긴급 환자에 대한 효율적인 위급상황 대처 효율을 높이는 효과를 가져올 수 있을 것이다. 또한 이러한 시스템은 구급차 내에 설치하여 환자 이송 중에 흉부 X-ray를 촬영, 초기진단이 가능하도록 하며, 질병의 감염부위 특정, 시간 단축, 환자 관리 시스템, 증강데이터 사용한 개선된 중증도 분류 기능을 통해 기존 x-ray 시스템과의 차별성을 두게 된다.

Abstract

Due to the global pandemic, the number of emergency patients continues to increase. More than half (56%) of patients visiting the emergency room are undergoing x-ray imaging. However, the time required for chest x-ray diagnosis in the emergency room takes about an hour, which is very long for emergency patients who need proper emergency treatment within golden time. In the case of system in this paper, x-ray imaging, which omits additional tests, is shortened to two stages of diagnosis, and the diagnosis speed of diagnosis is drastically shortened to around 12 seconds. To implement this, Classification and Detection-based Web Service was implemented, and Synthetic Medical X-ray Data were created using PGGAN to improve Classification performance. It is expected that these system will be used to improve the efficiency of medical process such as specifying problematic area, reducing x-ray processing time, and accurate disease classification in the emergency case.

Key words

AI Healthcare, Computer Vision, Classification, Detection, GAN, Web Development

* 이화여자대학교 소프트웨어학부 컴퓨터공학과, rhn9088@ewhain.net, 0613sy@ewhain.net, hieonn@ewha.ac.kr

* These authors contributed equally to this work

I. 서론

매년 천만명이 넘는 환자들이 응급실을 방문하며 [1] 응급실 방문 환자의 반 이상(56%)이 x-ray [2] 촬영을 진행하고 있다[3]. 하지만 응급실에서 흉부 x-ray 진단을 내리기 위해서는 x-ray 촬영, 추가 검사 여부 결정, mri/ct 등의 추가검사 진행, 진단의 4단계 과정을 거쳐야 하며 소요되는 시간은 약 1시간이다.[4] 이는 골든타임 내에 적절한 응급 치료를 하는 것이 중요한 응급환자에게는 긴 시간이다. 따라서 이를 단축하기 위해 다양한 시도들이 존재해 왔으며 본 논문에서는 이러한 문제를 해결하기 위해 x-ray 촬영만으로 질병 진단이 가능한 인공지능 진단 시스템을 설계하였다.

저장한다. 이러한 과정을 통해 최종적으로 웹사이트 화면의 환자의 질병 진단 결과와 질병 발병 의심부위를 Detection한 결과 화면이 출력된다. 가능성이 있는 질병을 3가지 제시하여 이를 통해 의사가 질병을 신속하게 진단할 수 있도록 돕는다.



그림 1. 서비스 사용 전후의 진단 과정
Fig. 1. Diagnosis before and after using service

II. 본론

2.1 시스템 구조

본 논문에서 개발한 시스템의 디러닝, 백엔드, 프론트엔드의 구성도는 <그림 2>와 같다. 우측부터 시작하여 웹사이트에서 x-ray 이미지를 업로드하면 백엔드에서 이미지의 Classification과 Detection 진단을 요청한다. 그 다음 학습된 모델로 업로드한 이미지를 진단(Classification)한다. Detection의 경우 yolov5 모델을 사용하여 학습을 진행한다. 그 다음 업로드된 이미지로 질병 의심 영역을 검출하고 결과 이미지를

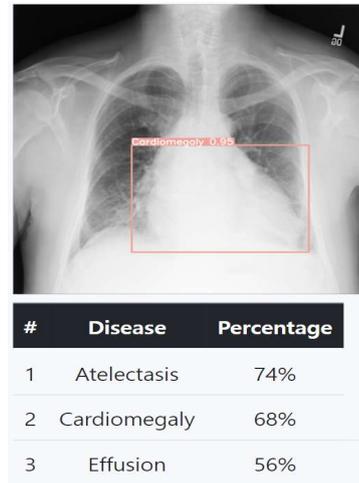


그림 3. 진단 결과
Fig. 3. Result of diagnosis

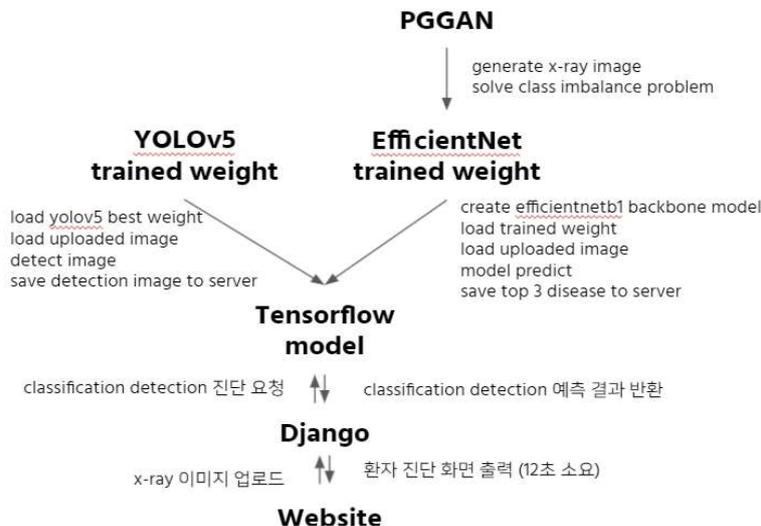


그림 2. 시스템 구조
Fig. 2. System structure

2.2 설계 및 구현

2.2.1 데이터셋

본 논문에서는 모델을 구축하기 위해 NIH(미국 국립 보건원) 제공 112,120개의 chest x-ray image 데이터[5] (이는 classification, detection에 훈련한다), 3,578개의 ChestX-Det-Dataset[6] (이는 detection에 훈련한다) 를 이용하였다. 진단하는 14가지 폐질환은 cardiomegaly, emphysema, effusion, hernia, infiltration, mass, nodule, atelectasis, pneumothorax, pleural_thickening, pneumonia, fibrosis, edema, consolidation 이다.

2.2.2 13가지 질병을 진단하는 Multi Label Classification 신경망 구축

13개의 폐질환을 진단하기 위해 Multi Label Classification 모델 훈련을 진행한다. 이때 해당 데이터셋에 가장 적합한 Backbone 모델을 찾기 위하여 Test Data AUC를 비교기준으로 6개의 Backbone 모델의 학습을 진행한다. 이 때 사용한 Backbone 모델로는 EfficientNet[7], MobileNet[8], Nasnet201[9], DenseNet201[10], InceptionResnetv2[11], Xception[12]가 있다. 학습 과정에서는 Backbone 모델들에 Pre-trained weight를 전이학습한다. 또한 LearningRateScheduler과 Early Stop Callback을 통해 오버피팅이 되는 것을 방지한다. 데이터의 질병 분포도를 유지하기 위해서는 MultilabelStratifiedKfold를 사용하며 같은 환자의 데이터가 중복되지 않도록 한다. 훈련된 가중치는 이 후 백엔드에서 사용하기 위해 저장한다. <그림 4>은 실험 결과로 출력된 Test dataset의 각 질병별 AUC값과 평균을 나타낸 것이다. 평균 AUC값이 0.7848로 가장 높은 EfficientNet[7]을 채택하였다.

2.3 질병 의심 영역을 검출하는 Bounding Detection

X-ray 진단이 더욱 신속하게 이루어지도록 질병 의심 영역을 검출한다. 본 논문에서는 Detection 모델 훈련을 위한 Backbone 모델로 YOLOv5[13][14][15] 를 사용한다.

총 14개의 질병을 Multi Label Detection을 한 결과 이미지는 <그림 5>와 같으며 훈련 mAP 결과는 0.97이다. <그림 6>은 이의 Metric 결과이다.

Disease/Model	EfficientNet	MobileNet	Nasnet201	DenseNet201	InceptionResNetv2	Xception
Cardiomegaly	0.838	0.844	0.741	0.745	0.775	0.776
Emphysema	0.87	0.867	0.813	0.855	0.821	0.834
Effusion	0.867	0.851	0.811	0.824	0.83	0.821
Hernia	0.65	0.67	0.78	0.686	0.708	0.799
Infiltration	0.702	0.683	0.614	0.678	0.647	0.648
Mass	0.804	0.786	0.681	0.776	0.789	0.78
Nodule	0.691	0.689	0.644	0.692	0.676	0.683
Atelectasis	0.784	0.762	0.711	0.759	0.723	0.727
Pneumothorax	0.851	0.832	0.808	0.834	0.813	0.834
Pleural Thickening	0.757	0.712	0.645	0.711	0.715	0.731
Pneumonia	0.691	0.686	0.594	0.701	0.671	0.692
Fibrosis	0.764	0.757	0.712	0.751	0.709	0.724
Edema	0.899	0.877	0.78	0.825	0.784	0.853
Consolidation	0.819	0.792	0.679	0.788	0.757	0.739
Average	0.7848	0.7691	0.7152	0.7575	0.7441	0.7586

그림 4. 6가지 모델에 대한 질병 AUC값
Fig. 4. AUC values from 6 models

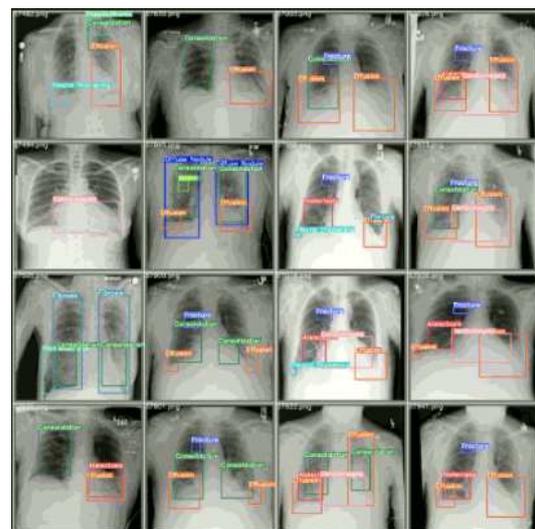


그림 5. YOLOv5를 이용한 질병 의심영역 검출 결과
Fig. 5. Results of disease detection using YOLOv5

epoch	299
train/box_loss	0.031153
train/obj_loss	0.017158
train/cls_loss	0.0099333
metrics/precision	0.89438
metrics/recall	0.93603
metrics/mAP_0.5	0.97203
metrics/mAP_0.5:0.95	0.70248
val/box_loss	0.018904
val/obj_loss	0.0077954
val/cls_loss	0.0031467
x/lr0	0.0020009
x/lr1	0.0020009

그림 6. Yolov5를 이용한 Detection 학습 결과
Fig. 6. Results of disease detection training using Yolov5

2.4 훈련 데이터셋의 Class Imbalance 완화를 위한 데이터 증강

의료데이터에서 흔히 발생하는 Class Imbalance를 해결하기 위해 PGGAN[16]을 이용, Synthetic Data를 생성한다. <그림 7>은 PGGAN을 이용하여 생성한 Chest X-ray 데이터이다.

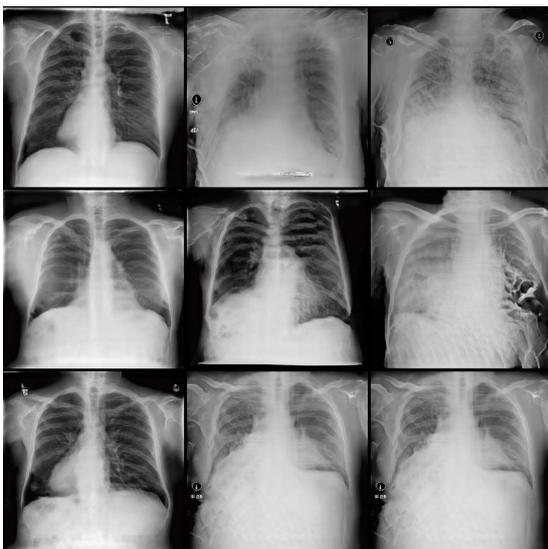


그림 7. PGGAN을 이용해서 생성한 X-ray데이터
Fig. 7. Generated image using PGGAN

이렇게 생성된 데이터를 사용하여 데이터셋의 Class Imbalance를 완화한 결과 질병 진단의 최종 평균 AUC 가 0.7848에서 0.8412로 0.056 향상되었다[17]. 또한 최종 Test Accuracy의 경우 Class Data Imabance 개선 전 0.969 에서 0.971로 개선되었다.

2.5 훈련된 모델의 이용

훈련된 모델은 Django를 활용하여 개발된 웹사이트에 API 형태로 구현하여, 폐질환 진단 예측 결과를 확률로 표시하여 결과를 돌려준다. 서버는 GPU를 탑재하지 않은 웹서버를 이용했을 때, 12초내에 결과를 돌려줄 수 있다. <표 1>은 국내 x-ray 타업체와의 진단소요시간과 정확도의 비교를 나타낸 표이다.

표 1. 진단소요시간과 정확도의 비교
Table. 1. Comparison of diagnosis time and accuracy

	진단소요시간	정확도
A사[18]	15초	97%
시스템	12초	97.1%

III. 결 론

3.1 기존 진단 서비스와의 차별성

현재 상용화되어 있는 AI기반 X-ray 진단 시스템에 비하여, 본 논문이 가지는 3가지의 차별성을 다음과 같다. 먼저, i5 CPU 8코어, NVIDIA GeForce MX250 GPU, 8G RAM 서버 환경에서도 12초 내의 빠른 속도 내에 진단이 가능하다. 두번째로는 중증도 분류 기능을 제공한다는 점이다. 마지막으로는 GAN 연구를 통해 모델의 정확도 향상을 시도했다는 점에서 기존의 인공지능 진단 솔루션과의 차별성을 가진다고 볼 수 있다.

3.2 시스템의 보건의료적 가치

우선, 본 논문의 시스템을 통해 한정된 시간 안에 더 많은 환자를 진단하여 응급환자에 대한 대처를 효율적으로 할 수 있다.

또한 위급한 응급상황에서 발생할 수 있는 진단 오류를 줄일 수 있어 전체적인 진단의 안정성을 향상시킬 수 있다. 마지막으로, 신속한 진단을 통해 추가 검사 및 대기 환자 관리에 소요되는 비용을 절감할 수 있다.

3.3 연구 성과 응용

본 논문의 시스템은 i5 CPU 8코어, NVIDIA GeForce MX250 GPU, 8G RAM 서버 정도인 가벼운 노트북 환경에서도 빠른 속도로 진단을 내릴 수 있다는 장점이 있다. 이를 응용하여 해당 서비스를 구급차에 설치하는 시나리오가 가능해진다. 휴대용 X-ray장비를 활용하여 구급차 내에서 X-ray 촬영 및 진단이 가능해진다면 더욱 빠르게 환자를 적합한 병원으로 이송할 수도 있다. 이를 수행하는 과정은 다음과 같다. 먼저, 휴대용 X-ray촬영 장비로 환자의 X-ray를 촬영하고 이를 구급차 내의 컴퓨터에 전송하여 진단을 내린다. 그 다음, 이러한 진단 결과를 바탕으로 가장 적합한 근처 병원 응급실 3곳에 환자의 X-ray사진과 진단 결과를 전송한다. 마지막으로 의사가 해당 결과를 확인하며 환자의 수용여부를 결정하여 구급차에 전송한다. 해당 병원에서 환자를 수용하기로 결정하면 환자를 최종적으로 해당 병원으로 이송한다. 이러한 과정을 통해 최종적으로 구급차 내에서 환자의 폐질환 진단이 가능해진다.

참 고 문 헌

- [1] 2019 Emergency Medical Statistical Yearbook: https://www.e-gen.or.kr/nemc/statistics_annual_report.do?brdclscd=01. [accessed: Nov. 01, 2021]
- [2] X-ray(Radiography)-Chest: <https://www.radiologyinfo.org/en/info/chestrad>. [accessed: Mar. 12, 2021]
- [3] B. J. Ahn, "Radiographic Status of the Visited Patients at University Hospital Emergency Room", Journal of the Korean Society of Radiology, Vol. 5, No. 2, pp. 81-92, 2011.
- [4] Lunit AI finds 10 lung diseases... "Like a video reading expert: <https://lunit-kr.prezly.com/lunis-ai-10gaji-pyejilhwan-chaj-anaendayeongsang-pandog-jeonmung-a-sujun>. [accessed: Oct. 10, 2021]
- [5] X. Wang, "ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases", arXiv 1705.02315, Vol. 5, No.1, 2017
- [6] Deepwise-AILab/ChestX-Det-Dataset: <https://github.com/Deepwise-AILab/ChestX-Det-Dataset>. [accessed: Nov. 28, 2021]
- [7] M. Tan, and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks", International Conference on Machine Learning, PMLR97, pp. 6105-6114, 2019.
- [8] G. Howard, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications", arXiv 1704.04861, Vol. 1, No. 1, 2017
- [9] Quoc V. Le, "Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition", CVPR, pp. 8697-8710, 2018
- [10] G. Huang, Z. Liu, L. Maaten, K. Q. Weinberger, "Densely Connected Convolutional Networks", CVPR, Vol. 5, No.1, 2017
- [11] K. He, X. Zhang, S. Ren, J. Sun, "Deep Residual Learning for Image Recognition", CVPR, Vol. 1, No. 1, pp. 770-778, 2016
- [12] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions", CVPR, Vol. 3, No. 1, pp. 1251-1258, 2017
- [13] M. Tan, Q. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks", arXiv:1905.11946, Vol. 5, No. 1, pp. 6105-6114, 2019
- [14] ultralytics/yolov5: <https://github.com/ultralytics/yolov5>. [accessed: Nov. 28, 2021]
- [15] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, "You only look once: Unified, real-time object detection", CVPR, pp. 779-788, 2016
- [16] A. Beers, J. Brown, K. Chang, J. P. Campbell, S. Ostmo, M. F. Chiang, J. Kalpathy-Cramer, "High-resolution medical image synthesis using progressively grown generative adversarial networks", CVPR, Vol.

2, 2018

- [17] H. B. Chung, S. Y. Choi, H. Kim, "A study on the impact of relieving the imbalance in data distribution between classes using PGGAN synthetic medical X-ray data on X-ray disease diagnostic classification accuracy", 2021
- [18] Lunit: <https://www.lunit.io/ko>. [accessed: Nov. 07, 2021]