

Diagnostic Classification of Chest X-ray Pneumonia using Inception V3 Modeling

Ji-Yul Kim¹, Soo-Young Ye^{2,*}

¹Daewoo general hospital

²Department of Radiology Catholic University of the Pusan

Received: October 17, 2020. Revised: November 27, 2020. Accepted: November 30, 2020

ABSTRACT

With the development of the 4th industrial, research is being conducted to prevent diseases and reduce damage in various fields of science and technology such as medicine, health, and bio. As a result, artificial intelligence technology has been introduced and researched for image analysis of radiological examinations. In this paper, we will directly apply a deep learning model for classification and detection of pneumonia using chest X-ray images, and evaluate whether the deep learning model of the Inception series is a useful model for detecting pneumonia. As the experimental material, a chest X-ray image data set provided and shared free of charge by Kaggle was used, and out of the total 3,470 chest X-ray image data, it was classified into 1,870 training data sets, 1,100 validation data sets, and 500 test data sets. I did. As a result of the experiment, the result of metric evaluation of the Inception V3 deep learning model was 94.80% for accuracy, 97.24% for precision, 94.00% for recall, and 95.59 for F1 score. In addition, the accuracy of the final epoch for Inception V3 deep learning modeling was 94.91% for learning modeling and 89.68% for verification modeling for pneumonia detection and classification of chest X-ray images. For the evaluation of the loss function value, the learning modeling was 1.127% and the validation modeling was 4.603%. As a result, it was evaluated that the Inception V3 deep learning model is a very excellent deep learning model in extracting and classifying features of chest image data, and its learning state is also very good. As a result of matrix accuracy evaluation for test modeling, the accuracy of 96% for normal chest X-ray image data and 97% for pneumonia chest X-ray image data was proven. The deep learning model of the Inception series is considered to be a useful deep learning model for classification of chest diseases, and it is expected that it can also play an auxiliary role of human resources, so it is considered that it will be a solution to the problem of insufficient medical personnel. In the future, this study is expected to be presented as basic data for similar studies in the case of similar studies on the diagnosis of pneumonia using deep learning.

Keywords: Pneumonia, Artificial Intelligence, Deep learning, Pneumonia Detection, Chest X-ray image

I. INTRODUCTION

4차 산업의 발전으로 의학·보건·바이오 등 여러 과학기술 분야에서는 질병을 예방하고 질병에 대한 피해를 줄이기 위하여 연구가 이루어지고 있다^[1]. 인공지능은 다양한 분야에서 연구 개발되고 있으며, 그중에서도 '사물 인식' 분야에서는 인간의 이미지 인식 정확도를 추월하는 수준에 이르렀다

[2].

폐렴은 호흡기 질환 중 사망자 수 1위 (대한민국 통계청, 2017)에 해당하는 질병이며 흉부 엑스레이 촬영은 폐렴을 진단하기 위한 일차적인 검사 방법이다^[3]. 흉부 엑스레이 촬영 이미지에서 병변을 찾아내기 위해서는 숙련된 의사가 직접 사진을 보고 판단해야 한다^[4]. 하지만 흉부 엑스레이 촬영 검사에서 폐렴은 다른 양성 이상을 나타내거나 해부학

* Corresponding Author: Soo-young Ye

E-mail: syye@cup.ac.kr

Tel: +82-51-510-0586

X선 영상 데이터 세트 (License: CC BY 4.0)를 사용하였다. 데이터 세트에는 전체 3,370개의 흉부 X선 영상 jpeg 이미지 파일로 구성되어 있다.

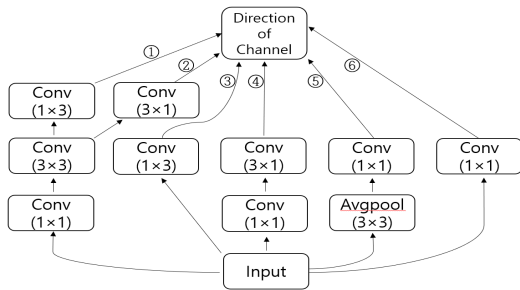


Fig. 2. The Inception module.

2. 실험방법

흉부 X선 영상을 이용한 폐렴의 분류와 검출에 대한 성능평가를 위하여 Inception V3 딥러닝 모델을 선정하였다. 그리고 흉부 X선 영상 파일을 데이터 세트로 하여 Inception V3 딥러닝 모델링을 구현하였다.

2.1 탐색적 데이터 분석

탐색적 데이터 분석은 다양한 시각으로 데이터를 관찰함으로써 데이터의 속성과 구조적 관계를 파악하고 이를 도표나 그래프로 나타내어 통계적, 직관적으로 데이터를 이해하는 과정이다^[3]. 본 연구에서는 전체 3,370개에 대한 탐색적 데이터 분석을 시행하여 흉부 X선 영상 데이터 중 정상 흉부 X선 영상의 경우 라벨링 0, 폐렴 흉부 X선 영상의 경우 라벨링 1로 라벨링 작업을 완료하였다.

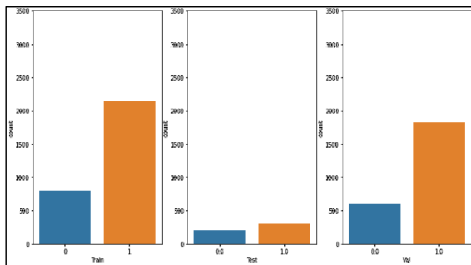


Fig. 3. The bar graph showing the number of labeling of the dataset.

2.2 데이터세트의 분류

캐글(Kaggle)에서 무료로 제공 및 공유하는 흉부

X선 영상 데이터 세트를 사용하였다. 전체 3,470개의 흉부 X선 영상 데이터 중 학습 데이터 세트 1,870개, 검증 데이터 세트 1,100개, 테스트 데이터 세트 500개로 분류하였다. 학습 데이터 세트의 경우 전체 1,870개의 흉부 X선 영상 데이터 중 라벨링 0으로 분류된 정상 흉부 X선 영상은 870개, 라벨링 1로 분류된 폐렴 흉부 X선 영상은 1,000개로 분류하였다. 검증 데이터 세트의 경우 전체 1,100개의 흉부 X선 영상 데이터 중 라벨링 0으로 분류된 정상 흉부 X선 영상은 500개, 라벨링 1로 분류된 폐렴 흉부 X선 영상은 600개로 분류하였다. 그리고 테스트 데이터 세트의 경우 전체 500개의 흉부 X선 영상 데이터 중 라벨링 0으로 분류된 정상 흉부 X선 영상은 200개, 라벨링 1로 분류된 폐렴 흉부 X선 영상은 300개로 분류하였다(Table 1).

Table 1. Classification of dataset

	Normal Chest image	Pneumonia Chest image
Training dataset	870	1,000
Validation dataset	500	600
Test dataset	200	300

2.3 딥러닝 모델링(Deep learning modeling)

Inception V3 딥러닝 모델의 기본 구조에 드롭아웃 층의 확률 p 값은 0.5를 적용하였으며, 활성화함수는 Sigmoid를 적용하였다. 그리고, 최적화 알고리즘은 Adam을 적용하였으며 학습 속도는 0.001을 적용하였다. 배치 정규화 알고리즘을 적용하여 배치 사이즈는 32로 설정하였으며, 학습 에포크는 35로 설정하여 흉부 X선 영상의 폐렴 검출 및 분류에 대한 딥러닝 모델링을 구현하였다.

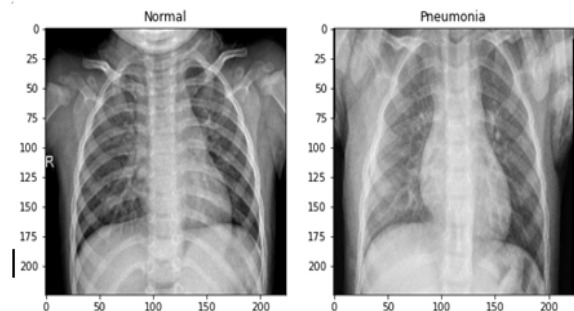


Fig. 4. Train dataset chest X-ray image.

3. 평가방법

흉부 X선 영상의 폐렴 검출 및 분류에 대한 딥러닝 모델링을 구현하였으며, 딥러닝 모델링의 에포크마다 출력되는 결과를 저장한 후 딥러닝 모델의 성능평가에 이용하였다. 성능평가 항목으로는 정확도(Accuracy), 정밀도(Precision), 재현율(Recall), F1 score, AUC(Area under the curve) score, 학습 모델링 및 검증 모델링의 정확도(Train & Validation accuracy), 손실함수 값(Train & Validation loss), 테스트 모델링의 Matrix 평가를 평가항목으로 선정하였다.

3.1 정확도 · 정밀도 · 재현율 · F 1 score · AUC score 평가

딥러닝 모델의 테스트 모델링의 결과 값인 예측값과 실측값의 대조 시의 정확한 정도를 나타내는 것이 정확도이며, 예측한 것 중 실측과 일치하는 비율을 나타내는 정도는 정밀도이다. 그리고 재현율이란 관련 있는 것으로 분류된 항목 중 실제 검색된 항목들의 비율을 의미하며, F1 스코어는 정밀도와 재현율의 평균을 의미한다.

AUC는 ROC(rec-eiver operating characteristic) curve의 아래 면적으로 데이터의 결과값이 우수할 수록 면적은 넓어지며 1에 가까운 값을 가진다^[12].

3.2 학습 모델링 및 검증 모델링 성능평가

학습 모델링 및 검증 모델링 과정의 결과 출력값인, 학습 정확도 학습 손실함수 값, 검증 정확도, 검증 손실함수 값을 구하여 딥러닝 모델의 성능 및 인공신경망의 학습상태를 평가하였다.

식 (1)은 정확도를 구하는 수식^[14]이다.

$$Accuracy = \frac{TN + TP}{TN + FP + FN + TP} \quad (1)$$

- T : 실제값
- F : 예측값
- N : 음성
- P : 양성

그리고 식 (2)는 손실함수로 가장 많이 쓰이는 평균 제곱 오차를 구하는 식^[14]이다.

$$E = \frac{1}{2} \sum_k (y_k - t_k)^2 \quad (2)$$

- y_k : 신경망의 출력
- t_k : 정답레이블
- k : 데이터의 차원 수

3.3 테스트 모델링 매트릭스 평가

테스트 데이터 세트를 이용하여 테스트 모델링을 수행한 후 정상 흉부 X선 영상 데이터와 폐렴 흉부 X선 영상 데이터의 실측결과와 출력된 예측결과를 매트릭스로 평가하였다.

III. RESULT

본 논문에서는 정상으로 판단되는 흉부 X선 영상 데이터세트와 폐렴으로 판단되는 흉부 X선 영상 데이터 세트를 통해 Inception V3 딥러닝 모델링을 이용한 폐렴 유무에 따른 분류 및 정확도를 평가하고자 하였다.

1. Metric 평가

1.1 정확도 · 정밀도 · 재현율 · F 1 score · AUC score

테스트 모델링의 결과값인 예측값과 실측값의 비교 평가를 위하여 정확도, 정밀도, 재현율, F1 score, AUC score를 구하여 평가하였다. Table. 2는 딥러닝 모델의 Metric 평가에 대한 결과값이다. Metric 평가의 결과에서 정확도는 94.80%, 정밀도는 97.24%, 재현율은 94.00%, F1 스코어는 95.59의 결과값을 나타내었다. 이러한 결과로 Inception V3 딥러닝 모델을 이용한 정상 흉부 X선 영상과 폐렴 흉부 X선 영상에 대한 분류 및 검출에 있어 성능이 매우 우수하다고 판단하였다.

Table 2. Metric results

Metric	Results
Accracy	94.80(%)
Precision	97.24(%)
Recall	94.00(%)
F1-score	95.59
AUC score	0.994

2. 학습 모델링 성능 평가 - Accuracy 평가

2.1 정확도 평가

흉부 X선 영상의 폐렴 검출 및 분류에 대하여 Inception V3 딥러닝 모델링에 대한 정확도를 평가하였다. Fig. 6.은 학습 모델링과 검증 모델링의 폐렴 유무에 따른 분류 및 정확도 평가를 도식화 한 그래프이다. 딥러닝 모델의 정확도는 최종 에포크 35에서 학습 모델링의 경우 94.91%, 검증 모델링은 89.68%의 정확도를 나타내었다. 이러한 결과로 Inception V3 딥러닝 모델은 흉부영상 데이터의 특징 추출 및 분류에 있어 매우 우수한 딥러닝 모델이라고 평가하였다.

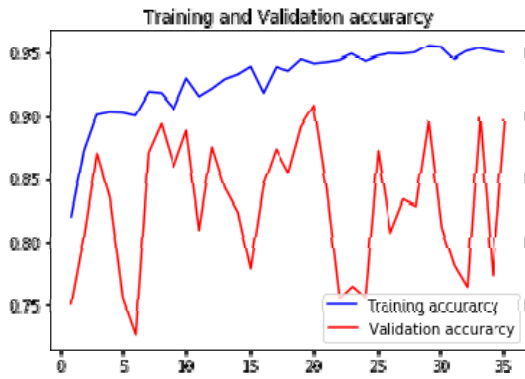


Fig. 6. Training and Validation accuracy.

판단되는데, ROC curve를 통해서도 AUC score의 값이 0.994인 것을 확인할 수 있다.

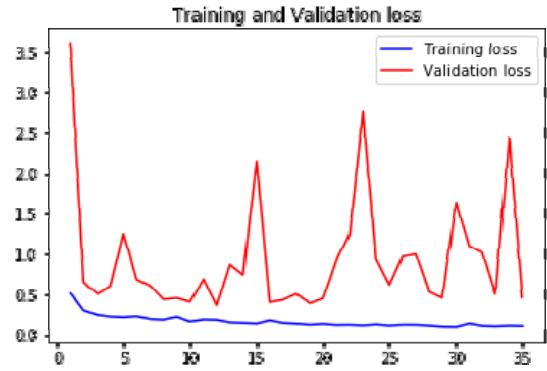


Fig. 7. Training and Validation loss.

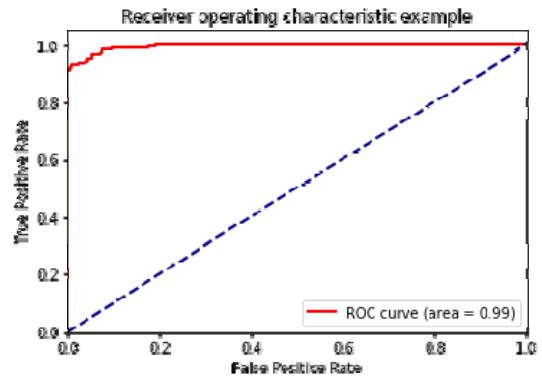


Fig. 8. Receiver operating characteristic curve.

2.2 손실함수 값 평가

흉부 X선 영상의 폐렴 검출 및 분류에 대하여 Inception V3 딥러닝 모델링에 대한 손실함수 값을 평가하였다. Fig. 7.은 학습 모델링과 검증 모델링의 손실함수 값을 도식화 한 그래프이다. 최종 에포크 35에서 손실함수 값의 평가는 학습 모델링은 1.127%, 검증 모델링은 4.603%의 손실함수 값을 나타내었다. 이러한 결과로 Inception V3 딥러닝 모델은 흉부영상 데이터의 특징 추출 및 분류를 위한 인공신경망의 학습상태가 우수하다고 평가하였다.

2.3 딥러닝 모델의 성능 평가 결과 - ROC curve

Fig. 8.는 ROC curve를 도식화 한 그래프이다. AUC score의 값이 1에 수렴 할수록 성능이 좋다고

3. 테스트 모델링 매트릭스 정확도 평가

흉부 X선 영상의 폐렴 검출 및 분류에 대하여 Inception V3 테스트 모델링을 평가하였다. Fig. 9.는 테스트 모델링의 흉부촬영 영상 폐렴의 유무를 실측값 및 예측값으로 평가한 매트릭스 정확도 평가를 도식화 한 그래프이다. 매트릭스 정확도 평가 결과 정상 흉부 X선 영상 데이터의 경우 전체 200건 중 192건은 정상, 1건은 폐렴으로 판단하여 96%의 정확도를 나타내었다. 폐렴 흉부 X선 영상 데이터의 경우 전체 300건 중 282건은 폐렴, 18건은 정상으로 판단하여 97%의 정확도가 입증되었다.

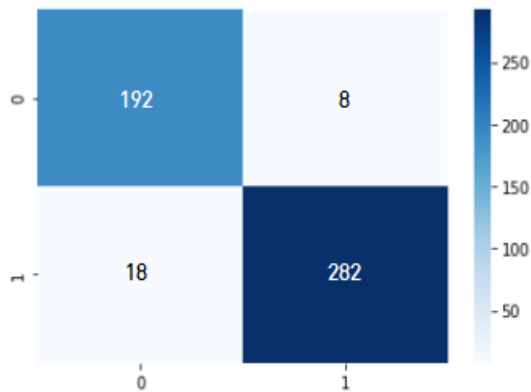


Fig. 9. Test modeling confusion matrix evaluation.

IV. CONCLUSION

본 논문에서는 흉부 X선 영상을 데이터셋으로 파이썬과 컨볼루션 인공신경망을 적용한 Inception V3 딥러닝 모델링을 구현하였다. 그리고 딥러닝 모델링을 적용한 폐렴의 분류와 검출에 있어 성능과 정확도를 평가하고자 하였다. 실험결과 Inception V3 딥러닝 모델은 성능평가 및 정확도 평가에서 폐렴의 분류 유무에 대하여 우수한 성능을 나타내었다. 이러한 결과로 Inception 계열의 딥러닝 모델의 경우 흉부 질환의 분류에 있어 유용한 딥러닝 모델이 될 것이라고 사료된다. 그리고 의료 인력의 보조적인 역할 또한 수행할 수 있을 것이라고 기대되어 부족한 의료인력 문제에도 해결점이 될 것이라고 사료된다. 향후 딥러닝을 이용한 폐렴의 진단에 대한 유사 연구 시 본 연구는 유사 연구의 기초 자료로 제시될 것이라고 기대된다.

Reference

[1] Becky McCall, "COVID-19 and artificial intelligence: protecting health-care workers and curbing the spread", *Lancet Digital Health*, Vol 2, No. s, pp. 166-167, 2020. [http://dx.doi.org/10.1016/S2589-7500\(20\)30054-6](http://dx.doi.org/10.1016/S2589-7500(20)30054-6)

[2] J. Y. Baek, "Recent artificial intelligence project trends, Institute for Information & communication Technology Planning & evaluation", *Weekly Technology Trend*, Vol. 1899, pp. 13-26, 2016.

[3] E. Musalar, S. Ekinçi, O. Ünek, E. Arş, H. Ş. Eren, B. Gürses, C. Aktaş, "Conventional vs invert-grayscale X-ray for diagnosis of pneumothorax in the emergency setting", *The American Journal of Emergency Medicine*, Vol. 35, No. 9, pp. 1217-1221, 2017. <http://dx.doi.org/10.1016/j.ajem.2017.03.031>

[4] Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning", arXiv:1711.05525, 2017.

[5] J. U. Sin, "Social Efforts for Loss and Recovery in the Medical Industry Following Coronavirus Infection-19", *Korea Institute for Health and Social Affairs*, Vol. 40, No. 3, pp. 1-6, 2020.

[6] Y. Heo, "Environmental Health Aspects on the Transmission of Coronavirus Infectious Disease-19", *Journal of Environmental Health Science*, Vol. 46, No. 1, pp. 1-2, 2020.

[7] M. Y. Cha, "Diagnosis and prediction of coronavirus-19 with artificial intelligence", *Corona-19 Science Report*, Vol. 4, No. 1, pp. 52-55, 2020.

[8] C. Y. Park, "AI Research Trends Looking Through 2018 ICML", *KAKAO AI report*, Vol. 14, No. 1, pp. 126-136, 2018.

[9] S. Chen, H. Wang, "SAR Target Recognition Based on Deep learning", *Data Science and Advanced Analytics*, Vol. 2, No. 1, pp. 541-547, 2014. <http://dx.doi.org/10.1109/DSAA.2014.7058124>

[10] Y. M. Seo, J. K. Han, "Deep Learning Algorithm to Identify Cancer Pictures", *The Korean Institute of Broadcast and Media Engineers*, Vol. 23, No. 5, pp. 669-681, 2018. <http://dx.doi.org/10.5909/JBE.2018.23.5.669>

[11] 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.

[12] 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>

- [13] J. E. Oh, S. W. Shin, G.G. Kim., "The role of computer-assisted diagnosis in 2D medical imaging", The Korean Institute of Electrical Engineers, Vol. 64, No. 7, pp. 16-25, 2015.
- [14] M. J. Warrens, "Cohen's kappa can always be increased and decreased by combining categories", Statistical Methodology, Vol. 7, No. 6, pp. 673-677, 2010.

Inception V3를 이용한 흉부촬영 X선 영상의 폐렴 진단 분류

김지율¹, 예수영^{2,*}

¹대우병원 영상의학과

²부산가톨릭대학교 방사선학과

요 약

4차 산업의 발전으로 의학·보건·바이오 등 여러 과학기술 분야에서는 질병을 예방하고 질병에 대한 피해를 줄이기 위한 연구가 이루어지고 있으며, 최근에는 ICT 기술의 발전과 더불어 인공지능 기술이 급부상하고 그 효용성이 입증되면서 영상의학 검사의 영상 분석에 인공지능 기술이 도입되어 연구되고 있다. 본 논문에서는 흉부 X선 영상을 이용하여 폐렴의 분류와 검출에 대한 딥러닝 모델을 직접 적용해보고 실제로 Inception 계열의 딥러닝 모델이 폐렴 검출에 있어 유용한 모델인지 평가하고자 한다. 실험재료는 캐글(Kaggle)에서 무료로 제공 및 공유하는 흉부 X선 영상 데이터 세트를 사용하였으며 전체 3,470개의 흉부 X선 영상 데이터 중 학습 데이터 세트 1,870개, 검증 데이터 세트 1,100개, 테스트 데이터 세트 500개로 분류하였다. 실험결과 Inception V3 딥러닝 모델의 Metric 평가에 대한 결과값은 정확도는 94.80%, 정밀도는 97.24%, 재현율은 94.00%, F1 스코어는 95.59의 결과값을 나타내었다. 그리고 흉부 X선 영상의 폐렴 검출 및 분류에 대하여 Inception V3 딥러닝 모델링에 대한 최종 에포크의 정확도는 학습 모델링의 경우 94.91%, 검증 모델링은 89.68%의 정확도를 나타내었다. 손실함수 값의 평가는 학습 모델링은 1.127%, 검증 모델링은 4.603%의 손실함수 값을 나타내었다. 이러한 결과로 Inception V3 딥러닝 모델은 흉부영상 데이터의 특징 추출 및 분류에 있어 매우 우수한 딥러닝 모델이며 학습상태 또한 매우 우수하다고 평가하였다. 테스트 모델링에 대한 매트릭스 정확도 평가 결과 정상 흉부 X선 영상 데이터의 경우 96%, 폐렴 흉부 X선 영상 데이터의 경우 97%의 정확도가 입증되었다. Inception 계열의 딥러닝 모델의 경우 흉부 질환의 분류에 있어 유용한 딥러닝 모델이 될 것이라고 판단되며 인력의 보조적인 역할 또한 수행할 수 있을 것이라고 기대되어 부족한 의료인력 문제에도 해결점이 될 것이라고 사료된다. 향후 딥러닝을 이용한 폐렴의 진단에 대한 유사 연구 시 본 연구는 유사 연구의 기초자료로 제시될 것이라고 기대된다.

중심단어: 폐렴, 인공지능, 딥러닝, 폐렴 검출, 흉부 X선 촬영영상

연구자 정보 이력

	성명	소속	직위
(제1저자)	김지율	대우병원 영상의학과	방사선사
(교신저자)	예수영	부산가톨릭대학교 방사선학과	부교수