Journal of RADIATION INDUSTRY
Vol. 18, No. 3, pp. 177-182, 2024
https://doi.org/10.23042/radin.2024.18.3.177

Original article

폐렴 및 정상군 판별을 위한 딥러닝 모델 성능 비교연구: CNN, VUNO, LUNIT 모델 중심으로

이지현[1,*] · 예수영[2]

[1]부산가톨릭대학교 방사선학과, [2]부산가톨릭대학교 방사선학과

A Comparative Study of Deep Learning Models for Pneumonia Detection: CNN, VUNO, LUIT Models

Ji-Hyeon Lee[1,*] and Soo-Young Ye[2]

[1]Marykynoll Medical Center, 121 Junggu-ro, Jung-gu, Busan 48972, Republic of Korea
[2]Department of Radiological Science, Catholic University of Pusan, 57 Oryundae-ro, Geumjeong-gu, Busan 46252, Republic of Korea

ABSTRACT The purpose of this study is to develop a CNN based deep learning model that can effectively detect pneumonia by analyzing chest X-ray images of adults over the age of 20 and compare it with VUNO, LUNIT a commercialized AI model. The data of chest X-ray image was evaluate based on accuracy,precision,recall,F1 score, and AUC score. The CNN model recored an accuracy of 82%, precision 76%, recall 99%, F1 score 86%, and AUC score 0.7937. The VUNO model recordded an accuracy of 84%, precision 81%, recall 94%, F1 score 87%, and AUC score 0.8233. The LUNIT model recorded an accuracy of 77%, precision 72%, recall 96%, F1 score 83%, and AUC score 0.7436. As a result of the Confusion Matrix analysis, the CNN model showe FN (3), showing the highest recall rate (99%) in the diagnosis of pneumonia. The VUNO model showed excellent overall perfomance with high accuracy (84%) and AUC score (0.8233), and the LUNIT model showed high recall rate (96%) but the accuracy and precision showed relatively low results. This study will be able to provide basic data useful for the development of a pneumonia diagnosis system by comprehensively considers the perfomance of the medel is necessary to effectively discriminate between penumonia and normal groups.

Key words: Pneumonia diagnosis system, Chest X-ray image, CNN, VUNO, LUNIT

1. 서 론

폐렴은 세균, 바이러스, 곰팡이 등의 병원체에 의해 폐의 염증성 질환이다. 이 질병은 전 세계적으로 주요 사망 원인 중 하나로, 면역력이 약한 성인에게 치명적일 수 있다[1,2].

폐렴의 증상은 발열, 기침, 가래, 호흡곤란 등이 있으며, 심각한 경우 패혈증이나 급성 호흡곤란증후군(ARDS)으로 발전할 수 있다. 이러한 증상은 다른 호흡기 질환과 유사하여 진단이 어려울 수 있으며, 특히 성인의 경우 만성 질환과 겹쳐 나타날 수 있어 더욱 복잡하다. 따라서 빠르고 정확한 진단이 필수적이다[3,4].

흉부 X-ray 촬영은 폐렴 진단에 가장 널리 사용되는 영상기술로, 폐의 염증을 시각적으로 확인 할 수 있다. 그러나 X-ray 이미지를 정확하게 해석하는 것은 방사선 전문의의 경험에 의존하기 때문에, 자원이 부족한 지역에서 어려움을 발생할 수 있다. 또한, 방사선 전문의 간의 해석 차이로 인해 진단의 일관성이 떨어질 수 있다[5].

이러한 문제를 해결하기 위해 최근 컴퓨터 비전 및 딥러닝 기술이 의료 영상 분석에 도입되고 있다. 특히 Convolution Neural Network(CNN)는 이미지 분류 및 객체 탐지에서 뛰어난 성능을 보여주고 있으며, 의료 영상 데이터 분석 분야에도 그 가능성을 인정받고 있다. CNN 기반 딥러닝 모델은 대규모 의료 영상 데이터를 학습하여 높은 정확도로 질병을 진단할 수 있으며, 이는 진단 속도와 정확성을 크게 향상하게 시킬 수 있다[6].

이 연구의 목적은 만 20세 이상의 성인 흉부 X-ray 영상을 분

ISSN 1976-2402 http://www.ksri.kr/

*Corresponding author: Jihyeon Lee
Tel: +82-51-803-6931, E-mail: qowhdfks@naver.com

Received 22 July 2024 Revised 1 August 2024 Accepted 12 August 2024

석하여 폐렴을 효과적으로 탐지할 수 있는 CNN 기반 딥러닝 모델을 개발하고, 이를 상용화된 AI 모델인 VUNO, LUNIT 모델과 비교하는 것이다. 이를 통해 모델의 장단점을 평가하고 실제 의료 현장에서의 적용 가능성을 탐구하고자 한다.

2. 재료 및 방법

2.1. 재료

2.1.1. 데이터수집 및 전처리

2023년 1월 1일부터 2024년 5월 12일까지 부산소재 M 병원에서 흉부 X-ray 촬영을 진행한 수검자 중 만 20세 이상 성인을 연구 대상으로 선정하여 흉부 X-ray 이미지 데이터를 수집하였다. 영상 데이터는 M병원의 PACS (Picture Archiving and Communication System)에서 수집되었으며, 흉부 X-ray 검사를 시행한 모든 수검자의 개인정보는 익명화되었다. 라벨링 작업을 위한 영상판독은 네 명의 영상의학과 판독전문의가 수행하였다.

데이터 세트는 랜덤샘플링을 통해 학습데이터와 테스트데이터로 나누었으며, 각각의 클래스 비율이 7:3으로 유지되도록 하였다. 흉부X-ray 영상 데이터 중 정상 흉부의 경우 라벨링을 0, 폐렴 흉부의 경우 1로 라벨링 작업을 완료하였다.

본 연구에서는 전체 2,331개의 흉부 X선 영상 JPEG 이미지 파일을 수집하여 학습 데이터 세트의 경우 전체 1,800개의 흉부 X-ray 영상 중 라벨링 0으로 분류된 정상 흉부 영상 900개, 라벨링 1로 분류된 폐렴 흉부 영상 900개로 분류하였다. 테스트 데이터 세트의 경우, 전체 531개의 흉부 X-ray 영상 중 라벨링 0으로 분류된 정상 흉부 영상 231개, 라벨링 1로 분류된 폐렴 흉부 영상 300개로 분류하였다(Table 1).

2.1.2. 사용 장비

일반촬영 장비는 삼성헬스케어 GC85A로 patient size medium 기준의 관전압 125 kVp, 관전류 320 mA로 설정하여 실험을 진행하였다.

2.1.3. 합성곱 신경망(Convolution Neural Network, CNN)

CNN은 하나 또는 여러 개의 컨볼류션 층과 그 위에 올려진 일반적인 인공신경망 층들로 이루어져 있으며, 컨볼루션 층에서 전처리를 수행하는 구조를 가진 인공신경망이다[7]. 이미지 데이터

Table 1. Classification of the data set

	0 (Normal)	1 (Pneumonia)
Training set	900	900
Test set	231	300
Total	1131	1200

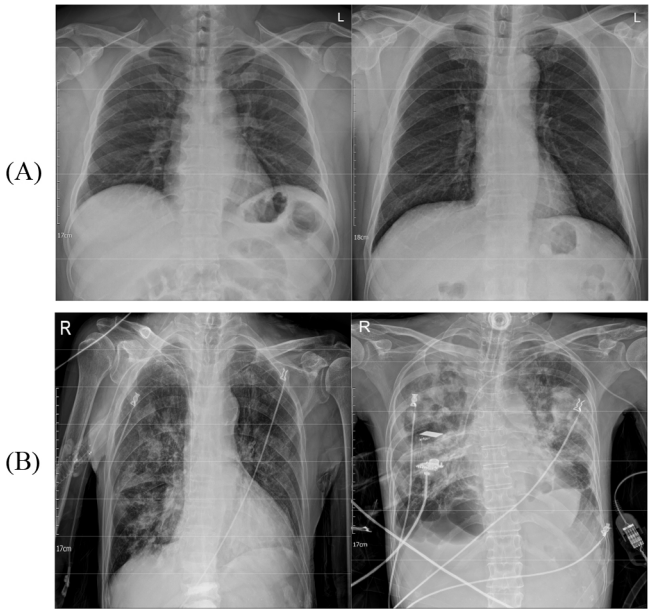

Fig. 1. Data samples from data set. (A) Normal cases (B) Pneumonia cases

의 특성을 효과적으로 학습하고, 이미지 분류, 객체 검출 등에서 뛰어난 성능을 보여, 의료 영상 분석 분야에서 널리 사용된다.

CNN은 세 가지 구성요소로 구성된다. 첫째, 컨볼류션 계층(Convolution layer)은 이미지의 공간적 계층적 특징을 학습하며, 작은 필터를 사용해 입력 이미지에서 특징맵을 생성한다. 둘째, 풀링 계층(Pooling layer)은 특징맵의 공간적 크기를 줄이고 연산량을 줄여 모델의 과적합을 방지한다. 셋째, 완전 연결 계층(Fully connected layer)은 추출된 특징을 기반으로 입력 이미지의 최종 분류 작업을 수행한다[8].

본 연구에서는 이러한 여러 개의 컨볼루션층와 풀링층, 마지막으로 완전 연결층으로 구성된 CNN을 사용하여 흉부 X-ray 이미지를 분석해 정상군, 폐렴을 분류하였다.

2.2. 방법

2.2.1. 학습모델링

딥러닝 모델로 Matlab 2023b를 사용하여 CNN 모델을 구현하였다. 모델 학습은 확률적 경사 하강법 모멘텀(SGDM, Stochastic Gradient Descent with Momentum) 최적화 알고리즘을 사용하였다. 확률적 경사 하강법(SGD, Stochastic Gradient Descent)은 미니배치를 사용하여 손실 함수의 기울기를 따라 모델 파라미터를 갱신한다. '확률적'이라는 것은 무작위로 선택한 미니배치를 사용한다는 것을 의미한다. 모멘텀은 기울기의 방향을 따라 수렴 속도를 높인다. SGDM은 이 두 가지 기법을 결합하여 더욱 빠르고 안정적으로 최적화를 수행한다[9].

모든 입력 RGB 이미지를 Grayscale 이미지로 변환하여 3채널

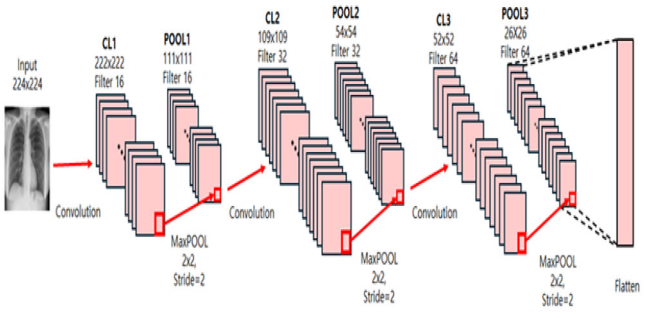

Fig. 2. Process of CNN.

Table 2. Training parameters of CNN model

Software	Matlab (2023b)
Input size	224×224
Optimization	SGDM
Epoch	5
Mini-Batch size	32
Learning rate	0.0001

에서 1채널로 줄임으로써 데이터양을 감소시키고 처리 속도를 향상 시켰다. 이후 성능 최적화와 범용적 적용 가능성을 높이기 위해 이미지를 동일한 크기인 224×224 픽셀로 조정하였다.

첫 번째 컨볼려션 레이어에 16개의 3×3 필터를 사용하고, 두 번째 레이어에는 32개, 세 번째 레이어에는 64개를 사용하였다. 활성화 함수는 ReLU (Rectified Linear Unit) 함수를 사용하여 비선형성을 추가하여 신경망 모델이 복잡한 패턴을 학습할 수 있도록 하였다. 2×2 풀링크기와 2의 스트라이드를 가진 최대풀링 (max pooling) 레이어를 거쳐 공간적 크기를 축소하였다.

CNN모델링의 학습은 총 5 에포크 동안 진행되었으며, 미니배치크기는 32, 학습률은 0.0001로 설정하여 학습 과정에서 모델 성능을 평가하였다.

2.3. 성능 평가

흉부 X선 영상의 폐렴 유무 분류에 관한 결과를 비교 평가하기 위해 각 에포크에서 생성된 결과를 기록한 후 CNN 모델의 성능 평가에 활용하였다. 학습된 모델을 검증 데이터로 평가하여 과적합 여부를 확인하고, 필요시 하이퍼파라미터를 조정하였다. 모델의 성능은 학습 모델링의 정확도 및 손실 함수, 테스트 모델링의 정확도 및 손실 함수값의 5회 모델링의 결과를 사용하여 평가하였다. 최종 모델은 테스트데이터를 사용하여 평가되었으며, 테스트 모델링에 대한 민감도, 특이도 지표를 통해 성능을 분석하였고, 폐렴 검출에 대한 모델의 성능을 기존 방법들과 비교하였다.

2.3.1. 정확도 평가

정확도 평가는 머신러닝 모델이 예측한 결과와 실제 정답이 얼마나 일치하는지를 나타내는 지표이다[10]. True Positive (TP)는 실제로 폐렴인 환자를 폐렴으로 정확하게 분류한 경우이다. True Negative (TN)는 실제로 정상인 환자를 정상으로 정확하게 분류한 경우이다. False Positive (FP)는 실제로 정상인 환자를 폐렴으로 잘못 분류한 경우이다. False Negative (FN)는 실제로 폐렴인 환자를 정상으로 잘못 분류한 경우이다.

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

TP: True Positive

TN: True Negative

FP: False Positive

FN: False Negative

2.3.2. 정밀도 평가

정밀도는 모델이 양성(폐렴)으로 예측한 사례 중 실제로 양성인 비율을 나타내는 성능 지표이다[10]. 이를 통해 모델이 양성으로 예측한 결과의 정확성을 평가할 수 있다.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

2.3.3. 재현율 평가

재현율은 실제 양성인 것 중에서 모델이 양성으로 예측한 비율을 나타내는 성능 지표이다[10]. 양성 클래스를 실제로 얼마나 잘 잡아내는지를 측정하여 모델의 감지 능력을 평가하는 중요한 지표이다.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

2.3.4. F1 점수

F1 점수는 정밀도와 재현율의 조화 평균으로 계산되는 성능 지표이다[10]. F1 점수는 정밀도와 재현율이 균형을 이루는지를 나타내며, 두 지표 모두 높은 값을 가질 때 상대적으로 높은 점수를 보인다. F1 점수가 높을수록 모델이 양성 클래스를 잘 예측함을 나타내며, 정밀도와 재현율이 균형을 이루고 있는지를 판단할 수 있다. 특히 데이터가 불균형할 때 유용하게 사용된다.

$$Flscore = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

2.3.5. ROC (Receiver Operating Characteristic) 및 AUC (Area Under Curve) 곡선

ROC 곡선은 여러 임계값들을 기준으로 FPR과 Recall의 변화를 시각화한 것이다[10]. Recall이 크고 FPR이 작을수록 성능이 좋다.

AUC는 ROC 곡선 아래의 면적을 의미한다[10]. AUC 값은 0과 1 사이의 값을 가지며, 모델의 성능을 한 숫자로 요약한다.

AUC가 1일 때 완벽한 분류기로 모든 양성 사례와 음성사례를 정확하게 예측한다.

2.3.6. 혼동행렬(Confusion Matrix)

혼동행렬은 모델링 매트리스 평가는 모델의 예측 결과를 표로 요약한 것으로, 예측 결과와 실제 라벨 간의 비교를 통해 모델의 정확도를 시각적으로 확인할 수 있다[10].

3. 결 과

본 연구에서는 흉부 X선 영상 데이터 세트를 이용하여 CNN 딥러닝을 통한 폐렴 유무 분류를 수행하고, 이를 상용화된 VUNO 모델 LUNIT 모델과 비교하였다.

3.1. Metric 평가

3.1.1. CNN 모델의 성능 평가

CNN 모델은 재현율이 0.99로 매우 높게 나타나, 실제 폐렴 환자를 대부분 정확히 검출하는데 탁월한 성능을 보였다. 그러나 정밀도는 0.76으로 높은 재현율에 비해 상대적으로 낮은 정밀도를 보였다.정확도는 0.82, F1점수는 0.86로, CNN 모델이 전반적으로 균형 잡힌 성능을 보인다는 것을 나타낸다. AUC score 0.7937은 다양한 임곗값에서 안정적으로 작동할 수 있음을 보여준다.

3.1.2. VUNO 모델의 성능 평가

VUNO 모델은 높은 정확도, 정밀도, F1점수, AUC점수를 기록하여 전반적으로 우수한 성능을 보였다. CNN 모델보다 정밀도와 정확도가 높아 폐렴이 아닌 경우를 더 정확히 예측할 수 있다. 재현율은 0.94로 높지만 CNN 모델보다는 다소 낮았다. VUNO 모델은 다양한 상황에서 폐렴을 안정적으로 폐렴을 진단할 수 있는 균형 잡힌 성능을 보여준다.

3.1.3. LUNIT 모델의 성능 평가

LUNIT 모델은 다른 두 모델에 비해 상대적으로 낮은 정확도와 정밀도를 보였다. 그러나 재현율이 0.96으로 높아 실제 폐렴 환자를 놓치지 않는 데 강점을 보인다. 그러나 AUC score가 가장 낮아, 예측의 일관성에는 다소 부족한 면이 있다.

Table 3. Construction of Deep learning modeling

Variavles	CNN	VUNO	LUNIT
Accuracy	0.82	0.84	0.77
Precision	0.76	0.81	0.72
Recall	0.99	0.94	0.96
F1 score	0.86	0.87	0.83
AUC score	0.7937	0.8233	0.7436

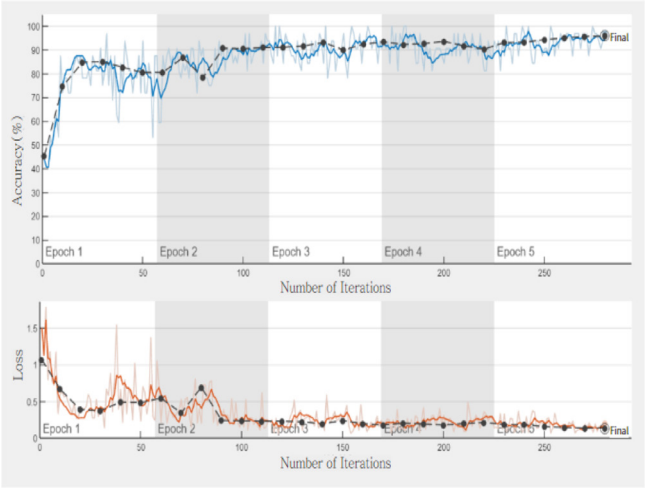

Fig. 3. Training accuracy and loss.

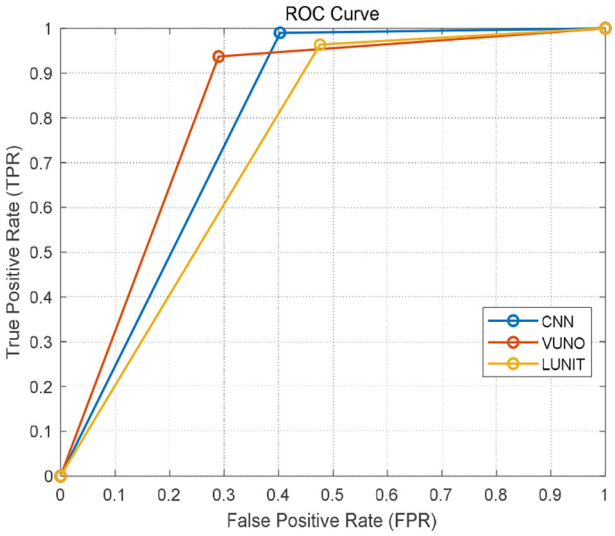

Fig. 4. Result of ROC curve.

3.2. 학습모델링 성능 평가

3.2.1. 손실 함수값 평가

흉부 X선 영상의 정상군과 폐렴 군을 분류하는 CNN모델의 손실함수값과 정확도를 평가하였다. Fig. 2는 학습과 검증모델링의 손실 함수값을 도식화한 그래프이다. 최종 에포크에서 학습 모델링의 손실함수 값은 0.2015, 검증 모델링의 손실함수 값은 0.1344로 나타났다. 이 결과로 보아 CNN 모델이 흉부 영상 데이터의 특징 추출 및 분류를 위한 인공신경망의 학습 상태가 우수하다고 평가할 수 있다.

2.3. 딥러닝 모델의 성능 평가-ROC curve, AUC score

ROC curve를 그리기 위해서 각 모델의 TPR (True Positive Rate)와 FPR (False Positive Rate)값이 필요하다. CNN은 TPR

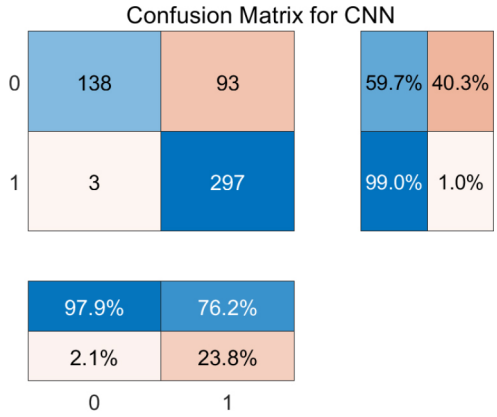

Fig. 5. Test modeling confusion matrix for CNN.

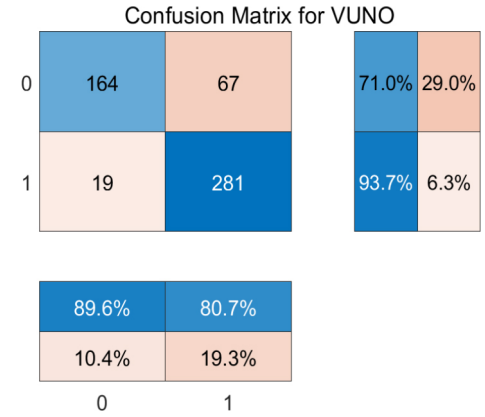

Fig. 6. Test modeling confusion matrix for VUNO.

0.99, FPR 0.40, AUC score 0.79이다. VUNO은 TPR 0.94, FPR 0.29, AUC score 0.82이다. LUNIT은 TPR 0.96, FPR 0.48, AUC score 0.74이다.

2.4. 테스트모델링의 confusion matrix 평가

Fig. 5는 CNN 모델을 사용한 테스트 모델링의 흉부 X선 영상의 폐렴 검출과 정상 분류에 대한 confusion matrix이다.

정상 흉부 X선 영상 데이터 전체 231건 중 138건은 정상, 93건은 폐렴으로 분류되어 59.7%의 정확도를 나타내었다. 폐렴 흉부 X선 영상 전체 300건 중 297건은 폐렴, 3건은 정상으로 분류되어 99%의 정확도를 나타내었다.

Fig. 6은 VUNO 모델을 사용한 테스트 모델링의 흉부 X선 영상의 폐렴 검출과 정상 분류에 대한 confusion matrix이다.

매트릭스 정확도 평가 결과 정상 흉부 X선 영상 전체 231건 중 164건은 정상, 67건은 폐렴으로 분류되어 71%의 정확도를 나타내었다. 폐렴 흉부 X선 영상 전체 300건 중 281건은 폐렴, 19건은 정상으로 분류되어 93%의 정확도를 나타내었다.

Fig. 7은 LUNIT 모델을 사용한 테스트 모델링의 흉부 X선 영상의 폐렴 검출과 정상 분류에 대한 confusion matrix이다.

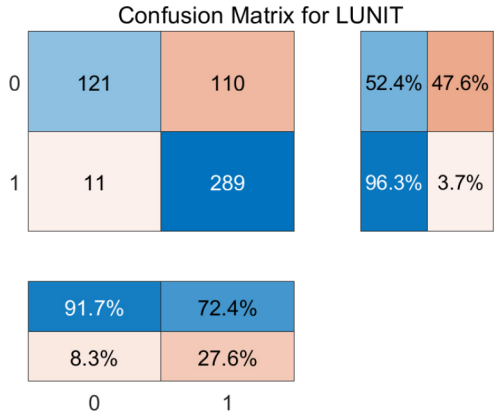

Fig. 7. Test modeling confusion matrix for LUNIT.

매트릭스 정확도 평가 결과 정상 흉부 X선 영상 데이터의 경우 전체 231건 중 121건은 정상, 110건은 폐렴으로 분류되어 52.4%의 정확도를 나타내었다. 폐렴 흉부 X선 영상 전체 300건 중 289건은 폐렴, 11건은 정상으로 분류되어 96.3%의 정확도를 나타내었다.

4. 결 론

본 연구에서는 CNN, VUNO, LUNIT 모델을 사용하여 흉부 X-ray 영상에서 폐렴을 진단하는 성능을 평가하였다. CNN 모델은 높은 재현율(Recall) 99%로 폐렴 환자를 대부분 정확하게 진단해 냈다. 하지만 False Positive 값이 93으로 높게 나타나 정상군을 폐렴으로 잘못 진단하는 경우가 있었다. 이는 CNN 모델이 폐렴 환자를 놓치지 않고 진단하는 데 강점이 있지만, 정밀도가 상대적으로 낮아 추가적인 검증이 필요함을 시사한다. VUNO 모델은 높은 정확도(84%)와 AUC score (0.8233)를 기록하여 전반적으로 우수한 성능을 보였다. LUNIT 모델은 높은 재현율(96%)을 보였으나 정확도(77%)와 정밀도(72%)에서 상대적으로 낮은 결과를 보였다.

CNN 모델은 VUNO, LUNIT 모델과 비교하여 유사한 수준의 성능을 보였다. 특히, 폐렴 환자를 정확하게 진단하는 능력에서는 가장 뛰어난 성능을 보이며, 폐렴 진단 시스템 개발에 있어 CNN 모델이 유용하게 활용될 수 있음을 보여준다. 앞으로의 연구에서는 이러한 모델의 성능을 더욱 개선하고, 다양한 임상 데이터를 통해 검증하여 더 신뢰성 있는 폐렴 진단 시스템을 구축하는 것이 중요하다. 본 연구는 폐렴 진단 시스템의 개발과 개선에 기초 자료로 활용될 수 있을 것이다.

5. 고 찰

폐렴 검출을 위한 딥러닝 모델인 VUNO 모델과 LUNIT 모델은 높은 민감도와 특이도를 가지고 있어 폐렴의 존재를 정확하게

감지할 수 있다. 그러나 이러한 성능은 폐렴 환자의 정확한 식별에 매우 유리하지만, 동시에 정상적인 경우에도 폐렴으로 오진할 가능성을 포함하고 있다.

특히, 높은 민감도는 정상일 때도 폐렴으로 오진할 가능성을 나타낸다. VUNO 모델은 FP 67건, LUNIT 모델은 110건으로, 정상일때도 폐렴으로 잘못 판단할 가능성을 나타낸다.

이러한 결과는 데이터 세트의 특성이나 모델의 학습 과정에서 발생하는 한계점이다. 따라서 폐렴 검출의 정확도를 높이기 위해서는 모델의 민감도와 특이도를 고려하여 최적의 임계값(threshold)을 설정하고, 추가적인 검토나 보정단계를 도입하여 모델의 신뢰성을 높이는 것이 중요하다. 실제 임상 환경에서의 적용 가능성을 높이기 위해 다양한 환자군과 상황에서의 모델 검증이 필수적이다. 이를 통해 폐렴 진단의 정확성을 높이고, 정상군의 오진을 최소화할 수 있을 것이다.

참고문헌

1. Tawsifur Rahman, Muhammad E.H. Chowdhury and Amith Khandaker. 2020. Transfer Learning with Deep Convolutional Neural Network (CNN) for Pneumonia Detection Using Chest X-ray. MDPI, Vol. 10, No. 3233, https://doi.org/10.3390/app10093233
2. Dalya S. AI-Dulaimi, Aseel Ghazi Mahmoud and Nadia Moqbel-Hassan. 2022. Development of Pneumonia Disease Detection Model Based on Deep Learning Algorithm. Wireless Communications and Mobile Computing, Vol. 2022, https://doi.org/10.1155/2022/2951168
3. Jaiswal A, Tuwari P and Kumar S. 2019. Identifying penumonia in chest X-rays: A deep learning approach. Elsevier, Vol. 145, pp 511-518. https://doi.org/10.1016/j.measurement.2019.05.076
4. Song HJ, Lee EB and Jo HJ. 2020. Evaluation of Classification and Accuracy in Chest X-ray images using Deep Learning with Convolution Neural Network. *Korean Soc. Radiol*. Vol 14, No 1, https://doi.org/10.7742/jksr.2019.14.1.39
5. Kim JH, Kim JY and Kim GH. 2020. Clinical Validation of a Deep Learning Algorighm for Detection of Pneumonia on Chest Radiographs in Emergency Department Patients with Acute Febrile Respiratory Illness. Vol. 9, No. 1981, https://doi.org/10.3390/jcm9061981
6. Ni Q, Sun ZY and Qi L. 2020. A deep learning approach to characterize 2019 coronavirus disease (COVID-19) pneumonia in chest CT images. *European Soc. Radiol.* **30**:6517-6527. https://doi.org/10.1007/s00330-020-07044-9
7. Kim JY and Ye SY. 2020. Diagnostic Classification of Chest X-ray Pneumonia using Inception V3 Modeling. *Korean Soc. Radiol*. **14**(6):773-780. https://doi.org/10.7742/jksr.2020.14.6.773
8. Kim JY and Ye SY. 2022. Comparative Evaluation of Chest Image Pneumonia based on Learning Rate Application. *Korean Soc. Radiol.* **16**(5):595-602. https://doi.org/10.7742/jksr.2022.16.5.595
9. Kang MJ. 2020. Comparison of Gradient Descent for Deep Learning. *Korean Academia-Industrial cooperation Soc*. 21(2):189-194. https://doi.org/10.5762/KAIS.2020.21.2.189
10. Kim JY and Ye SY. 2021. Comparative Analysis by Batch Size when Diagnosing Penumonia on Chest X-Ray Image using Xception Modeling. *Korean Soc. Radiol*. **16**(5):545-554. https://doi.org/10.7742/jksr.2021.15.4.547