

공학석사학위논문

CNN 모델을 활용한
흉부 X-ray 이미지 기반 폐렴 진단

Diagnosis of Pneumonia Bases on
Chest X-ray Images Using CNN Model

충북대학교 대학원

산업인공지능학과

김 정 호

2025년 6월

공학석사학위논문

CNN 모델을 활용한
흉부 X-ray 이미지 기반 폐렴 진단

Diagnosis of Pneumonia Bases on
Chest X-ray Images Using CNN Model

지도교수 류관희

산업인공지능학과

김 정 호

이 논문을 공학석사 학위논문으로 제출함.

2025년 6월

본 논문을 명성구의 공학석사학위 논문으로 인정함.

심 사 위 원 장 류 관 희 ①

심 사 위 원 박 태 형 ①

심 사 위 원 전 명 근 ①

충 북 대 학 교 대 학 원

2025년 6월

차 례

Abstract	ii
표 차례	iii
그림 차례	iv
I. 서 론	1
1. 연구의 동기 및 배경	1
2. 프로젝트의 목적	2
3. 프로젝트의 기대 효과	2
II. 이론적 배경	3
1. CNN 모델 개요	3
1.1. ChexNet 모델	4
1.2. ResNet-50 모델	6
1.3. EfficientNetB0 모델	7
2. 데이터 증강 및 전처리	9
2.1. 데이터 증강의 필요성	9
2.2. 대표적인 데이터 증강 기법	10
2.3. 데이터 전처리 과정	10
3. 의료 영상 분석에서 CNN 활용	12
3.1. CNN이 의료 영상 분석에 적합한 이유	12
3.2. 의료 영상 분석에서 CNN의 주요 응용 분야	13
3.2. 의료 영상 분석에서 CNN의 성능 평가	13

III. 연구 방법	15
1. 데이터셋	15
1.1. 데이터 출처	15
1.2. 데이터셋 구성	16
1.3. 데이터 전처리	17
2. CNN 모델 적용	18
2.1. 모델 아키텍처 설명	19
2.2. 하이퍼파라미터 설정	21
3. 데이터 증강 기법 적용	23
3.1. 회전(Rotation)	23
3.2. 확대 및 축소(Zooming)	24
3.3. 좌우 반전(Flipping)	24
3.4. 밝기 조절(Brightness Adjustment)	24
3.4. 잘림(Cropping)	24
4. 성능 평가 지표 및 방법	25
4.1. 정확도	25
4.2. 정밀도, 재현율, F1-score	26
4.3. AUC-ROC curve	27
IV. 실험 및 결과	28
1. 1차 실험: 모델 성능 비교	28
1.1. 학습데이터와 검증데이터 분리	28
1.2. CheXNet, ResNet-50, EfficientNetB0 성능 비교	29
1.3. Confusion Matrix 결과	31
1.4. 손실 및 정확도 그래프	33
2. 2차 실험: ChexNet 모델 하이퍼파라미터 튜닝	35
2.1. 하이퍼파라미터 조정: 학습률, 배치 사이즈, 에포크 수	35

2.2. 조정된 CheXNet 모델 성능 분석	36
2.3. Confusion Matrix 및 성능 평가	36
3. 최종 결과 분석 및 비교	40
3.1. CheXNet 모델의 최종 성능	40
3.2. 하이퍼파라미터 조정 전후 성능 비교	41
3.3. 모델별 성능 평가 요약	42
V. 결론	43
1. 결론	43
2. 한계점 및 발전방향	44
참고문헌	46

Diagnosis of Pneumonia Bases on Chest X-ray Images Using CNN Model

Kim, Jung Ho

*Department of Industrial AI
Graduate School, Chungbuk National University
Cheongju, Korea
Supervised by Professor Ryu, Kwan Hee*

Abstract

* A thesis for the degree of Master in August 2025.

표 차례

표 1. 페럼 이미지 데이터셋 분포표	17
표 2. 페럼 진단에 최적화된 CNN 모델 비교	21
표 3. 페럼 클래스에 대한 결과 및 정확도	29
표 4. 정상 클래스에 대한 결과	30
표 5. 하이퍼파라미터 조정 전후 성능 비교표	41
표 6. 모델별 성능 평가 요약	42

그림 차례

그림 1. CheXNet 모델의 흉부 엑스레이 히트맵	4
그림 2. CheXNet 모델 아키텍처	5
그림 3. ResNet-50 모델 아키텍처	7
그림 4. EfficientNetB0 모델 아키텍처	8
그림 5. EfficientNetB0 Confusion Matrix	31
그림 6. RestNet-50 Confusion Matrix	32
그림 7. CheXNet Confusion Matrix	32
그림 8. 정확도 그래프	34
그림 9. 손실 그래프	34
그림 10. Confusion Matrix	37
그림 11. 정확도 그래프	39
그림 12. 손실 그래프	39

I. 서 론

1. 연구의 동기 및 배경

폐렴은 전 세계적으로 사망률이 높은 질병 중 하나로, 특히 노인과 어린이, 그리고 면역력이 약한 사람들에게 큰 위험을 초래하는 질병이다. 전통적으로 폐렴 진단은 흉부 X-ray 이미지를 기반으로 한 방사선 전문의의 판독에 의존해 왔다. 하지만 X-ray 이미지를 해석하는 과정은 시간 소모적이며, 방사선 전문의의 경험과 주관적 판단에 따라 결과가 달라질 수 있어 일관된 진단을 보장하기 어렵다. 특히, 전문 인력이 부족한 지역에서는 이러한 문제로 인해 폐렴 진단이 지연되거나 오류가 발생할 수 있다.

인공지능(AI) 기술의 발전과 더불어, 의료 영상 처리 분야에서도 딥러닝 알고리즘을 활용한 자동화된 진단 시스템 개발이 활발히 이루어지고 있다. 특히, 컨볼루션 신경망(CNN)은 이미지 분석에 있어 탁월한 성능을 보이며, 의료 영상의 패턴 인식 및 질병 분류에서 중요한 역할을 하고 있다. CNN 모델은 대규모 데이터셋에서 사전 학습된 가중치를 바탕으로 특정 작업에 맞게 미세 조정하는 방식으로, 데이터가 제한적인 상황에서도 효과적인 성능을 낼 수 있어 흉부 X-ray 이미지 분석에 적합하다.

2. 프로젝트의 목적

본 연구의 목적은 CNN 모델을 활용하여 흉부 X-ray 이미지를 기반으로 폐렴을 자동으로 진단할 수 있는 시스템을 개발하는 것이다. 다양한 CNN 모델을 비교 분석하여, 가장 높은 성능을 보이는 모델을 선택하고 하이퍼파라미터 튜닝 및 데이터 증강을 통해 진단의 정확도를 향상시키는 데 중점을 둔다. 또한, 이 시스템이 의료진의 주관적 판단에 의한 오류를 줄이고, 진단의 일관성과 신속성을 보장할 수 있도록 하는데 기여하고자 한다.

3. 프로젝트의 기대효과

자동화된 폐렴 진단 시스템은 의료 현장에서 여러 이점을 제공할 수 있다. 첫째, 진단 과정을 자동화함으로써 진단 시간을 단축시킬 수 있다. 둘째, 인공지능 기반 시스템은 일정한 기준에 따라 진단을 내리기 때문에 의료진 간의 주관적인 차이를 줄여줄 수 있다. 셋째, 의료 인력이 부족한 지역이나 환경에서도 원격으로 폐렴 진단을 지원할 수 있는 가능성을 열어준다. 특히, 최근 COVID-19 팬데믹과 같은 전염병 상황에서 폐 질환의 신속한 진단과 대응이 중요한 만큼, 본 연구가 실질적인 임상 적용에 기여할 수 있을 것으로 기대된다.

II. 이론적 배경

1. CNN 모델 개요

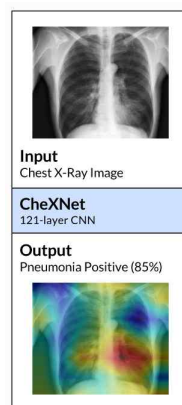
컨볼루션 신경망(CNN, Convolutional Neural Network)은 딥러닝에서 이미지 분석과 패턴 인식 작업에서 가장 널리 사용되는 모델입니다. CNN은 이미지 내의 특징을 자동으로 추출하고, 학습된 가중치를 통해 새로운 이미지에서 동일한 특징을 감지할 수 있는 능력을 갖추고 있습니다. 하지만 CNN을 처음부터 학습시키려면 대규모의 데이터셋과 충분한 학습 시간이 필요하며, 이는 자원 소모가 크고 실제 프로젝트에서 실용적이지 않은 경우가 많습니다. 이 문제를 해결하기 위해, 대규모 데이터셋에서 미리 학습된 CNN 모델을 사용하여 새로운 작업에 맞춰 모델을 미세 조정(fine-tuning)하는 전이 학습(Transfer Learning)이 개발되었습니다.

CNN 모델은 이러한 전이 학습 기법을 기반으로 하며, 대규모 데이터셋(예: ImageNet)에서 사전 학습된 가중치를 활용해 빠르게 성능을 향상시킬 수 있습니다. 특히 의료 영상 분석과 같은 복잡한 문제에서 충분한 학습 데이터가 부족하거나 학습 시간이 제한적인 상황에서 유용합니다. 본 연구에서는 흉부 X-ray 이미지 데이터를 분석해 폐렴을 진단

하기 위해 세 가지 주요 CNN 모델을 사용했습니다. 각각의 모델은 이미지 분류에서 높은 성능을 자랑하며, 이 연구의 목적에 맞게 적합한 성능을 발휘합니다. 다음은 CheXNet, ResNet-50, EfficientNetB0 모델에 대한 설명입니다.

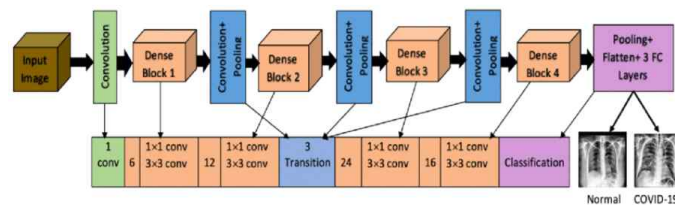
1.1. CheXNet 모델

CheXNet은 Stanford University에서 개발된 의료 영상 분석에 특화된 CNN 모델로, 주로 흉부 X-ray 이미지를 분석하여 폐렴을 진단하는데 사용됩니다. CheXNet은 DenseNet-121 모델을 기반으로 하며, 대규모 ImageNet 데이터셋으로 사전 학습된 가중치를 사용하여 폐렴 진단을 위한 전이 학습을 수행합니다. CheXNet은 흉부 X-ray 이미지에서 폐렴뿐만 아니라, 14가지 흉부 질환을 감지할 수 있는 모델로 설계되었으며, 이미지 분류와 패턴 인식 성능에서 탁월한 성과를 보였습니다.



<그림 1 - CheXNet 모델의 흉부 엑스레이 히트맵>

DenseNet 구조의 특성상, CheXNet은 이전 층에서 학습된 특징들이 이후 층으로 전달되어 더 풍부한 정보를 학습할 수 있는 장점을 가집니다. DenseNet은 각 레이어가 이전 레이어의 출력을 모두 입력으로 사용하므로, 정보의 손실을 최소화하고, 필요한 정보가 네트워크 전반에 걸쳐 고르게 전달될 수 있도록 합니다. 이 구조 덕분에 CheXNet은 더 적은 파라미터로도 높은 성능을 낼 수 있으며, 층이 깊어질수록 발생하는 기울기 소실(vanishing gradient) 문제도 완화할 수 있습니다.



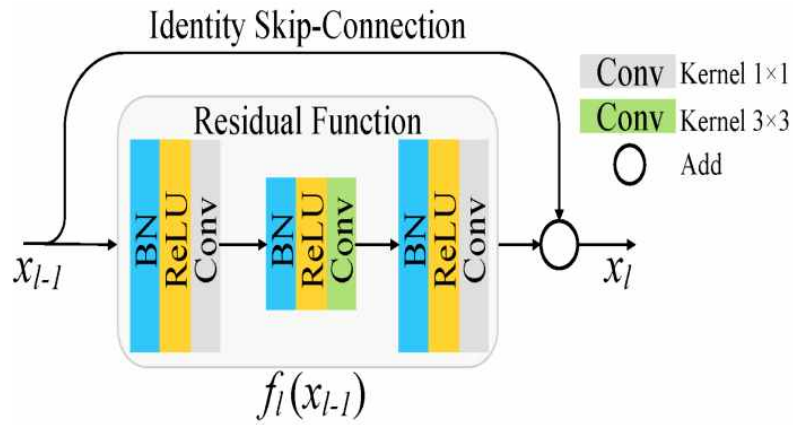
<그림2 - CheXNet 모델 아키텍처>

CheXNet은 2017년 발표된 이후, 흉부 X-ray 이미지 기반 폐 질환 진단에서 매우 높은 성능을 보여주었으며, 기존 방사선 전문의들의 진단 정확도와 비슷하거나 이를 뛰어넘는 성과를 냈습니다. 특히, 의료 영상에서의 활용도가 높아 다양한 임상 환경에서도 적용 가능성이 큼니다. 이 연구에서는 CheXNet이 제공하는 DenseNet-121 기반의 강력한 특징 학습 능력을 활용하여, 흉부 X-ray 이미지에서 폐렴을 진단하는 정확도를 높이는 것을 목표로 합니다.

1.2. ResNet-50 모델

ResNet(Residual Network)은 Microsoft에서 개발한 CNN 모델로, 매우 깊은 네트워크에서도 효과적으로 학습할 수 있는 잔차 학습(residual learning) 개념을 도입했습니다. 일반적으로 신경망의 층이 깊어질수록 기울기 소실(vanishing gradient) 문제로 인해 학습이 어려워지는 경향이 있습니다. 하지만 ResNet은 이러한 문제를 해결하기 위해 잔차 연결(shortcut connections)을 사용하여 네트워크의 성능을 크게 향상시켰습니다.

ResNet-50은 이러한 잔차 학습을 기반으로 한 50층의 깊은 신경망 구조를 가지고 있으며, 복잡한 이미지 데이터에서 높은 성능을 자랑합니다. 이 모델은 기본적으로 층이 깊어질수록 발생할 수 있는 학습 문제를 잔차 연결을 통해 완화하며, 비교적 간단한 구조로도 복잡한 패턴을 인식하는 능력이 뛰어납니다. 특히, ResNet-50은 다양한 이미지 분류 작업에서 매우 좋은 성능을 보였으며, 폐렴과 같은 질병 진단에서도 널리 사용되고 있습니다.



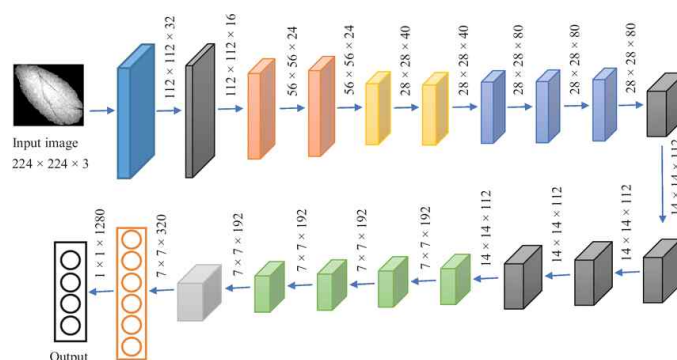
<그림3 - ResNet-50 모델 아키텍처>

ResNet의 장점은 단순히 깊은 네트워크만으로 성능을 향상시키는 것이 아니라, 층이 깊어지면서도 기울기 소실을 피하고 안정적인 학습을 유지할 수 있다는 점입니다. 이러한 이유로, ResNet-50은 흉부 X-ray 이미지 기반 폐렴 진단에서도 그 성능을 입증할 수 있었으며, 다른 모델들과 비교했을 때도 빠른 학습 속도와 높은 정확도를 보여주었습니다. 전이 학습을 통해 기존에 학습된 가중치를 적용하여 빠르고 효율적인 모델 학습이 가능하며, 이 연구에서도 ResNet-50을 활용하여 폐렴 진단 성능을 평가합니다.

1.3. EfficientNetB0 모델

EfficientNetB0는 Google에서 개발된 경량화된 CNN 모델로, 성능과 효율성을 극대화한 아키텍처를 가지고 있습니다. EfficientNet 모델은

복합 계수(compound scaling) 기법을 도입하여, 모델의 깊이(depth), 너비(width), 해상도(resolution)를 균형 있게 확장할 수 있는 구조를 가지고 있습니다. 이러한 확장 방식 덕분에, EfficientNet은 적은 파라미터로도 매우 높은 성능을 발휘할 수 있습니다. EfficientNetB0는 이 시리즈 중에서 가장 작은 모델이지만, 성능 면에서 여전히 매우 효율적입니다.



<그림4 - EfficientNetB0 모델 아키텍처>

EfficientNetB0는 ImageNet 데이터셋으로 사전 학습되었으며, 전이 학습을 통해 의료 영상 분석 작업에서도 좋은 성능을 발휘합니다. EfficientNetB0는 특히 파라미터 수가 적어 계산 자원 사용이 효율적이고, 빠른 추론 속도를 제공하면서도 높은 정확도를 보장하는 모델입니다.

2. 데이터 증강 및 전처리

딥러닝 모델의 성능은 학습에 사용되는 데이터의 양과 질에 크게 의존합니다. 특히, 의료 영상 분석과 같은 작업에서는 대규모의 고품질 데이터를 확보하기 어렵기 때문에 데이터 증강(data augmentation)과 전처리(preprocessing)가 필수적입니다. 데이터 증강은 기존의 데이터를 변형하여 새로운 데이터로 만들어내는 기술로, 모델의 일반화 성능을 향상시키고 과적합(overfitting)을 방지할 수 있습니다.

2.1 데이터 증강의 필요성

의료 영상 데이터는 그 수집 과정에서 제한이 있을 뿐만 아니라, 질병 진단에서 중요한 패턴이 특정 위치나 크기에 따라 다양하게 나타날 수 있습니다. 이 때문에, 제한된 양의 데이터로 모델을 학습시키는 것은 모델이 특정 패턴에 과적합하는 문제를 야기할 수 있습니다. 이를 해결하기 위해 데이터 증강 기법을 사용하여 원본 이미지를 다양한 방식으로 변형하고, 모델이 다양한 경우의 수를 학습할 수 있도록 돕습니다. 데이터 증강을 통해 학습 데이터의 다양성을 높이고, 모델이 보다 일반화된 성능을 보일 수 있도록 합니다.

2.2 대표적인 데이터 증강 기법

다양한 데이터 증강 기법이 있으며, 본 연구에서 사용한 주요 기법은 다음과 같습니다:

회전(Rotation): 이미지를 임의의 각도로 회전시켜 다양한 방향에서 패턴을 학습할 수 있도록 합니다.

확대 및 축소(Zooming): 이미지의 크기를 확대하거나 축소하여 모델이 다양한 크기의 패턴에 적응할 수 있도록 합니다.

좌우 반전(Flipping): 이미지를 좌우로 반전시켜, 대칭적인 패턴에 대해서도 학습할 수 있게 합니다.

밝기 조절(Brightness Adjustment): 이미지의 밝기를 조정하여, 모델이 다양한 조명 조건에서도 일관된 성능을 낼 수 있도록 합니다.

잘림(Cropping): 이미지의 일부분을 잘라내어, 부분적인 패턴을 학습할 수 있게 합니다.

2.3 데이터 전처리 과정

데이터 전처리는 모델 학습 전에 데이터를 정제하고, 일관된 입력 형식으로 변환하는 과정입니다. 의료 영상 데이터는 이미지 형식, 크기, 해상도 등이 다양하기 때문에, 모델에 적합한 형태로 전처리하는 것이 필수적입니다. 본 연구에서 사용된 주요 전처리 단계는 다음과 같습니다:

크기 조정(Resizing): 모든 X-ray 이미지를 동일한 크기(예: 224x224)

로 조정하여 모델의 입력 형식에 맞춥니다.

정규화(Normalization): 각 이미지의 픽셀 값을 0과 1 사이로 정규화하여, 학습 과정에서 특정 값에 대한 편향을 줄이고, 학습 속도를 높입니다.

채널 수 조정: 컬러 이미지의 경우 RGB 채널을 사용하지만, 흉부 X-ray 이미지는 흑백(단일 채널)이므로 필요한 경우 채널 수를 맞추는 작업이 필요합니다.

데이터 증강과 전처리 과정은 딥러닝 모델의 학습을 위해 필수적인 단계이며, 특히 의료 영상에서 그 중요성이 큼니다. 본 연구에서는 이러한 기법들을 통해 학습 데이터의 양과 질을 향상시킴으로써, 모델의 성능을 높이고, 과적합 문제를 줄였습니다.

3. 의료 영상 분석에서 CNN 활용

의료 영상 분석 분야에서 인공지능, 특히 딥러닝 기술의 활용이 활발히 이루어지고 있습니다. CNN은 그 중에서도 이미지 분류 및 패턴 인식에서 뛰어난 성능을 발휘하기 때문에 의료 영상 분석에 적합한 알고리즘으로 주목받고 있습니다. 의료 영상 분석에서는 방사선 사진, CT 스캔, MRI 등의 고차원 데이터에서 질병의 징후를 자동으로 감지하고 분류하는 작업이 필요하며, CNN은 이러한 작업에서 중요한 역할을 합니다.

3.1 CNN이 의료 영상 분석에 적합한 이유

의료 영상 분석에서 CNN의 성공적인 활용은 여러 가지 이유에서 비롯됩니다. 첫째, CNN은 자동으로 이미지에서 중요한 특징을 학습하고, 이 특징을 바탕으로 질병을 분류할 수 있습니다. 이는 기존의 수작업으로 특징을 추출하는 방식보다 훨씬 효율적이며, 이미지의 복잡한 구조에서 중요한 패턴을 감지하는 능력이 뛰어납니다.

둘째, CNN은 계층적 특징 학습(hierarchical feature learning)을 통해 이미지의 저수준 특징(예: 모서리, 텍스처)부터 고수준 특징(예: 특정 장기 구조의 이상)까지 다양한 수준에서 패턴을 학습할 수 있습니다. 이를 통해, CNN은 단순한 이미지 분류를 넘어서 복잡한 질병 징후까지 감지할 수 있게 됩니다.

3.2 의료 영상 분석에서 CNN의 주요 응용 분야

의료 영상에서 CNN의 주요 응용 분야는 다음과 같습니다:

질병 진단 및 분류: CNN은 흉부 X-ray, CT, MRI와 같은 영상 데이터를 바탕으로 암, 폐렴, 심장 질환 등의 질병을 자동으로 진단하고 분류하는 데 사용됩니다. 특히, CheXNet과 같은 모델은 폐렴을 포함한 여러 가지 흉부 질환을 높은 정확도로 진단할 수 있습니다.

세분화(Segmentation): CNN을 기반으로 한 모델은 단순히 질병을 분류하는 것 외에도, 장기나 병변 부위를 자동으로 분리해내는 세분화 작업에서도 활용됩니다. 이는 특정 장기에서 질병이 발생한 위치를 정확히 파악하는 데 중요한 역할을 합니다.

검출(Detection): 의료 영상에서 특정 병변이나 이상 징후를 자동으로 검출하는 작업에서 CNN이 사용됩니다. 예를 들어, 폐암의 경우 종양을 자동으로 검출하는 모델이 개발되었으며, 이는 전문 의료진의 진단을 보조하는 중요한 기술로 활용되고 있습니다.

3.3 의료 영상 분석에서 CNN의 성능 평가

CNN의 성능은 주로 정확도(Accuracy), 정밀도(Precision), 재현율(Recall), F1-score, AUC-ROC 곡선 등 다양한 지표를 통해 평가됩니다.

다. 이러한 지표들은 모델이 얼마나 잘 질병을 진단하고, 그 과정에서 오류를 최소화하는지 평가하는 데 사용됩니다. 특히 의료 분야에서는 거짓 음성(False Negative)의 비율을 최소화하는 것이 중요하며, 이로 인해 재현율이 중요한 평가 지표로 활용됩니다.

CNN은 의료 영상 분석에서 기존의 전통적인 방법보다 우수한 성능을 발휘하는 것으로 입증되었으며, 여러 연구에서 이를 활용한 자동화 시스템이 개발되고 있습니다. 특히, 흉부 X-ray 이미지를 통해 폐렴 진단에서 우수한 성능을 보이는 CheXNet, ResNet, EfficientNet과 같은 CNN 모델들은 의료 현장에서 활용 가능성이 매우 큽니다. 본 연구에서는 이러한 CNN 기반 모델을 활용하여 흉부 X-ray 이미지를 분석하고, 폐렴 진단의 정확도를 높이는 것을 목표로 합니다.

Ⅲ. 연구 방법

1. 데이터셋

본 연구에서 사용된 데이터셋은 폐렴 진단을 위해 흉부 X-ray 이미지를 포함하는 대규모 공개 데이터셋을 기반으로 하였습니다. 이 데이터셋은 CNN 모델이 폐렴과 정상 상태를 분류하는 데 필요한 충분한 양의 이미지를 제공하며, 전처리와 데이터 증강을 통해 학습에 적합하게 구성되었습니다. 데이터셋은 학습, 검증, 테스트 세트로 나누어져 있으며, 각 세트는 모델 성능을 평가하는 데 사용되었습니다.

1.1. 데이터 출처

사용된 흉부 X-ray 데이터는 Kaggle에서 제공하는 'Chest X-ray Images (Pneumonia)' 데이터셋을 기반으로 하였습니다. 이 데이터셋은 실제 의료 환경에서 수집된 X-ray 이미지로 구성되어 있으며, 폐렴에 걸린 환자의 X-ray 이미지뿐만 아니라 건강한 사람의 X-ray 이미지도 포함하고 있습니다.

이 데이터셋은 National Institutes of Health (NIH)의 연구를 바탕으로 만들어졌으며, 총 5,856장의 흉부 X-ray 이미지로 구성되어 있습니다. 각 이미지는 흑백(단일 채널) 이미지로, 2차원으로 표현된 흉부의 방사선 사진입니다. 이미지의 해상도는 각기 다르며, 모델 학습을 위해 일정한 크기로 조정되었습니다.

데이터는 다음과 같은 두 가지 주요 클래스(Class)로 분류됩니다:

정상(Normal): 건강한 상태의 흉부 X-ray 이미지

폐렴(Pneumonia): 폐렴이 진단된 상태의 흉부 X-ray 이미지

1.2. 데이터셋 구성

데이터셋은 학습, 검증, 테스트 세트로 나누어져 있으며, 각 세트는 모델 학습 및 평가에 사용됩니다.

Train Set: 모델 학습에 사용되는 데이터셋으로, 전체 이미지 중 70%에 해당하는 4,215장의 이미지로 구성되었습니다. 이 데이터셋은 정상 이미지와 폐렴 이미지가 포함되어 있으며, 모델이 학습하는 데 필요한 다양한 데이터를 제공합니다.

Validation Set: 학습 중 모델의 성능을 평가하고, 과적합을 방지하기 위해 사용되는 검증 데이터셋으로, 정상 이미지 8장, 폐렴 이미지 8장으로 총 16장으로 구성되었습니다.

Test Set: 최종적으로 학습된 모델의 성능을 평가하는 데 사용되며, 전체 데이터의 20%인 1,056장의 이미지로 구성되었습니다.

구체적인 데이터 분포는 아래와 같습니다:

Set	Class	Images
Train set	Normal	1341
Train set	Pneumonia	2874
Validation set	Normal	8
Validation set	Pneumonia	8
Test set	Normal	234
Test set	Pneumonia	390

<표1 - 폐렴이미지 데이터셋 분포표>

데이터셋에서 폐렴 환자의 이미지가 더 많은 비중을 차지하고 있어, 이는 클래스 불균형(class imbalance) 문제를 발생시킬 수 있습니다. 이를 해결하기 위해 데이터 증강을 적용하고, 클래스 가중치 조정을 통해 학습 중 모델이 특정 클래스에 치우치지 않도록 하였습니다.

1.3. 데이터 전처리

데이터셋의 이미지들은 다양한 크기와 해상도를 가지기 때문에, CNN 모델의 입력 형식에 맞게 일관된 크기로 조정되었습니다. 모든 이미지는 224x224 픽셀로 크기 조정(resizing)되었으며, 모델의 학습 속도를 높이고 과적합을 방지하기 위해 정규화(normalization)도 적용되었습니다. 이를 통해 각 이미지의 픽셀 값은 [0, 1] 사이의 값으로 변환되었으며, 이러한 정규화는 모델이 특정 값에 편향되지 않고 일관된 학습을 할 수 있도록 돕습니다.

또한, 데이터셋 내에 클래스 간 비율 차이가 존재하기 때문에, 데이터 증강 기법을 통해 학습 데이터셋의 다양성을 높였습니다. 이를 통해 폐렴과 정상 상태의 X-ray 이미지 모두에서 다양한 패턴을 학습할 수 있도록 하였습니다.

2. CNN 모델 적용

본 연구에서는 흉부 X-ray 이미지 기반 폐렴 진단을 위해 CNN 모델을 사용하여 빠르고 효율적인 학습을 진행했습니다. CNN 모델은 다양한 이미지 분류 작업에서 높은 성능을 보이며, 특히 전이 학습(transfer learning)을 통해 이미 사전 학습된 가중치를 활용할 수 있다는 점에서 의료 영상 분석에 적합합니다. 본 연구에서 사용된 주요 모델은 CheXNet, ResNet-50, EfficientNetB0이며, 이 모델들은 대규모 데이터셋인 ImageNet으로 미리 학습된 가중치를 바탕으로 흉부 X-ray 이미지에서 폐렴을 진단하는데 적용되었습니다.

2.1. 모델 아키텍처 설명

CheXNet (DenseNet-121 기반): CheXNet은 Stanford University에서 개발된 폐렴 진단에 특화된 CNN 모델로, DenseNet-121을 기반으로 구축되었습니다. DenseNet-121의 주요 특징은 밀집 연결(dense connections)입니다. 이는 각 층(layer)이 이전 모든 층의 출력을 입력으로 사용하여 정보 전달을 극대화하며, 층이 깊어지더라도 중요한 특징이 소실되지 않도록 보장합니다. DenseNet 구조 덕분에 CheXNet은 효율적인 학습이 가능하며, 적은 파라미터로도 높은 성능을 보입니다.

특징:

- DenseNet-121 구조를 바탕으로 CNN 레이어의 출력을 재사용하여 더 깊은 특징을 학습.
- ImageNet 데이터셋으로 사전 학습된 가중치를 사용하여 빠른 전이 학습이 가능.
- 폐렴 진단에 특화된 추가 데이터셋으로 미세 조정(fine-tuning)되어 흉부 X-ray 이미지에 최적화됨.

CheXNet은 DenseNet-121의 구조를 그대로 사용하면서도 흉부 X-ray 이미지에 최적화된 가중치와 손실 함수를 통해 폐렴을 진단하는 데 필요한 높은 분류 성능을 제공합니다. 이 모델은 대규모 의료 영상 데이터셋에 대해 검증되었으며, 높은 진단 정확도를 자랑합니다.

ResNet-50:

ResNet-50은 Microsoft에서 개발된 잔차 학습(Residual Learning)을 도입한 CNN 모델입니다. CNN의 층이 깊어질수록 발생하는 기울기 소실(vanishing gradient) 문제를 해결하기 위해, ResNet은 잔차 연결

(shortcut connections)을 사용하여 네트워크의 깊이를 확장하면서도 성능 저하를 방지합니다. ResNet-50은 50개의 레이어로 구성되어 있으며, 대규모 이미지 분류 작업에서 높은 성능을 발휘하는 모델입니다.

특징:

- 잔차 연결을 사용해 깊은 네트워크에서 발생하는 기울기 소실 문제를 해결.
- 50개의 레이어로 구성되어 복잡한 패턴을 학습할 수 있음.
- 전이 학습을 통해 기존의 사전 학습된 가중치를 활용하여 흉부 X-ray 이미지에서 폐렴을 진단.

ResNet-50의 장점은 단순히 깊은 네트워크를 갖추고 있을 뿐만 아니라, 잔차 연결을 통해 깊이 있는 학습이 가능하다는 점입니다. 이를 통해 폐렴과 같은 질병의 미세한 패턴도 효과적으로 학습할 수 있습니다.

EfficientNetB0:

EfficientNetB0는 Google에서 개발된 경량화된 CNN 모델로, 성능과 효율성을 극대화한 아키텍처를 가지고 있습니다. EfficientNet은 모델 크기를 늘리거나 줄일 때 균형을 맞추는 복합 계수(compound scaling) 기법을 도입하였으며, 이를 통해 모델의 깊이, 너비, 해상도를 균형 있게 확장할 수 있습니다. EfficientNetB0는 그 중에서도 가장 작은 규모의 모델로, 적은 파라미터로도 높은 성능을 발휘할 수 있습니다.

특징:

- EfficientNetB0는 적은 양의 연산 자원으로도 높은 정확도를 제공하는 모델로, 컴퓨팅 자원이 제한된 환경에서 효과적임.

- 복합 계수(compound scaling)를 통해 모델의 깊이, 너비, 해상도를 균형 있게 확장 가능.
- ImageNet 데이터셋으로 사전 학습되었으며, 전이 학습을 통해 의료 영상 분석 작업에서도 우수한 성능을 보임.

EfficientNetB0는 CheXNet과 ResNet-50에 비해 더 적은 파라미터로 구성되어 있지만, 매우 높은 효율성을 자랑하며, 의료 영상 분석과 같은 작업에서도 충분히 우수한 성능을 보일 수 있는 모델입니다. 특히, 적은 자원으로 빠른 추론을 필요로 하는 경우에 적합합니다.

모델	핵심 아키텍처	주요 특징	장점	폐렴진단 최적화 방법
CheXNet	DensNet-121	밀집 연결	<ul style="list-style-type: none"> - 효율적인 학습 - 적은 파라미터로 높은 성능 - 높은 정확도 	폐렴 데이터셋에 대한 미세 조정으로 높은 성능 제공
RestNet-50	전차 학습	전차 연결을 통한 기울기 소실 방지	<ul style="list-style-type: none"> - 복잡한 패턴 학습 가능 	전이 학습을 통해 폐렴 진단
EfficientNetB0	EfficientNet	복합 스케일링	<ul style="list-style-type: none"> - 적은 자원으로 높은 효율성 - 제한된 자원 환경에 적합 	전이 학습을 통해 폐렴 진단

<표2 - 폐렴 진단에 최적화된 CNN 모델 비교>

2.2. 하이퍼파라미터 설정

모델 학습에서 하이퍼파라미터 설정은 모델 성능에 중요한 영향을 미칩니다. 하이퍼파라미터는 학습 과정에서 고정된 값으로, 모델이 학습을 얼마나 빠르고 정확하게 진행할 수 있는지를 결정합니다. 본 연구에서는 CheXNet 모델을 중심으로 하이퍼파라미터를 조정하였으며,

ResNet-50과 EfficientNetB0에도 유사한 설정을 적용하였습니다.

- 배치 사이즈(Batch Size): 배치 사이즈는 학습할 때 한 번에 처리하는 데이터 샘플의 수를 의미합니다. 본 연구에서는 32로 설정하였습니다. 배치 사이즈가 너무 크면 메모리 사용량이 증가하고, 너무 작으면 학습이 불안정해질 수 있습니다. 32는 모델이 안정적으로 학습할 수 있는 적절한 값입니다.

- 학습률(Learning Rate): 학습률은 가중치가 업데이트되는 속도를 결정하는 중요한 하이퍼파라미터입니다. 너무 크면 모델이 빠르게 수렴하지 못하고, 너무 작으면 학습 속도가 느려질 수 있습니다. 본 연구에서는 0.0001의 학습률을 사용하여 모델이 안정적으로 수렴할 수 있도록 하였습니다.

- 에포크 수(Epochs): 에포크 수는 전체 데이터셋을 몇 번 반복해서 학습할지를 결정합니다. 본 연구에서는 50에포크로 설정하였으며, 이는 모델이 충분히 학습하면서도 과적합되지 않도록 조정한 값입니다. 학습 과정 중 성능 향상이 미미해지면 학습을 조기에 종료하는 방식(early stopping)도 도입하여, 불필요한 학습을 방지하였습니다.

- 옵티마이저(Optimizer): 옵티마이저는 가중치 업데이트 방식을 결정하는 알고리즘입니다. 본 연구에서는 Adam 옵티마이저를 사용하였습니다. Adam은 학습률을 자동으로 조정하면서 빠른 학습 속도와 안정적인 수렴을 제공합니다.

- 손실 함수(Loss Function): 손실 함수는 모델의 예측과 실제 값 사이

의 차이를 계산하는 함수입니다. 본 연구에서는 Binary Crossentropy를 손실 함수로 사용하였습니다. 이는 이진 분류 문제에서 일반적으로 사용되는 손실 함수로, 모델이 폐렴 여부를 정확히 분류할 수 있도록 합니다.

이러한 하이퍼파라미터 설정을 통해 모델은 안정적이고 빠르게 학습할 수 있었으며, 각 모델의 성능이 최적화될 수 있었습니다.

3. 데이터 증강 기법 적용

데이터 증강(data augmentation)은 모델이 다양한 입력 데이터를 학습할 수 있도록 인위적으로 데이터를 변형하여 새로운 샘플을 생성하는 기법입니다. 의료 영상 분석에서는 충분한 양의 학습 데이터 확보가 어려운 경우가 많고, 그로 인해 모델이 과적합(overfitting)되는 문제가 발생할 수 있습니다. 특히, 흉부 X-ray 이미지와 같은 의료 데이터는 수집이 어려우므로, 데이터 증강 기법을 활용해 기존 데이터를 변형하여 모델이 더 다양한 패턴을 학습할 수 있도록 해야 합니다.

본 연구에서는 다음과 같은 데이터 증강 기법을 적용하여 학습 데이터를 확장하고, 모델의 일반화 성능을 향상시켰습니다.

3.1 회전(Rotation)

이미지를 임의의 각도로 회전시키는 기법을 사용하여, 다양한 방향에서 폐렴과 관련된 패턴을 학습할 수 있도록 했습니다. 본 연구에서는

이미지를 $\pm 15^\circ$ 이내로 회전시켜, 실제 촬영 상황에서 발생할 수 있는 각도 변화를 학습할 수 있게 했습니다.

3.2 확대 및 축소(Zooming)

이미지의 크기를 확대하거나 축소하는 기법을 통해, 모델이 다양한 크기에서의 패턴을 인식할 수 있도록 학습했습니다. 확대 및 축소는 이미지에서 특징이 더 명확하게 드러날 수 있는 경우와, 작게 나타날 수 있는 경우를 모두 고려해 모델이 유연하게 대처할 수 있게 해줍니다.

3.3 좌우 반전(Flipping)

이미지를 좌우로 반전시키는 기법을 적용하여, 좌우 방향에서 발생할 수 있는 पैрам 패턴을 학습할 수 있도록 했습니다. 의료 영상의 경우 특정 방향성에 의존하지 않는 것이 중요하므로, 반전을 통해 다양한 시각 패턴을 모델이 학습할 수 있도록 합니다.

3.4 밝기 조절(Brightness Adjustment)

X-ray 이미지는 촬영 환경에 따라 밝기 변화가 발생할 수 있습니다. 이를 고려하여, 본 연구에서는 이미지의 밝기를 조정하는 데이터 증강 기법을 적용하였습니다. 이를 통해 모델이 다양한 조명 조건에서도 일관된 성능을 낼 수 있도록 했습니다.

3.5 잘림(Cropping)

이미지의 일부를 잘라내는 기법을 통해 부분적으로 나타나는 특징을 학습하게 했습니다. 이는 실제 임상 환경에서 이미지가 전체적으로 나타나지 않는 경우에도 모델이 일관된 성능을 낼 수 있도록 돕습니다.

데이터 증강은 본 연구에서 중요한 역할을 하였으며, 이를 통해 모델이 다양한 상황에서 폐렴 진단을 정확하게 할 수 있도록 하였습니다. 데이터 증강을 통해 기존의 데이터셋을 확장함으로써, 더 많은 패턴을 학습하고 모델의 일반화 능력을 높였습니다.

4. 성능 평가 지표 및 방법

딥러닝 모델의 성능을 평가하는 것은 연구의 핵심입니다. 본 연구에서는 CNN 모델의 성능을 평가하기 위해 다양한 평가 지표를 사용했습니다. 이 지표들은 모델이 폐렴과 정상 상태를 얼마나 잘 분류하는지를 다양한 측면에서 평가하며, 의료 영상 분석에서는 정확하고 신뢰할 수 있는 모델 성능이 매우 중요합니다.

4.1. 정확도(Accuracy)

정확도(Accuracy)는 모델이 전체 데이터 중 얼마나 많은 샘플을 올바르게 분류했는지를 나타내는 지표입니다. 이는 전체 샘플에서 올바르게 분류된 샘플의 비율로 계산됩니다. 본 연구에서 정확도는 다음과 같이 계산되었습니다:

$$\text{정확도} = \frac{\text{올바르게 분류된 샘플 수}}{\text{전체 샘플 수}}$$

정확도는 모델의 전반적인 성능을 평가하는 데 유용하지만, 클래스 불균형 문제가 있는 데이터셋에서는 정확도가 과대 평가될 수 있습니다. 본 연구의 경우, 폐렴과 정상 데이터의 비율이 차이가 있기 때문에 정확도 외에도 다른 성능 지표를 함께 사용하였습니다.

4.2. 정밀도(Precision), 재현율(Recall), F1-score

정밀도(Precision)는 모델이 양성으로 예측한 샘플 중에서 실제로 양성인 샘플의 비율을 나타냅니다. 본 연구에서는 폐렴을 양성(positive), 정상 상태를 음성(negative)으로 간주하였으며, 정밀도는 모델이 폐렴으로 진단한 환자 중 실제로 폐렴인 환자의 비율을 평가합니다.

$$\text{정밀도} = \frac{\text{참 양성 (True Positive)}}{\text{참 양성 (True Positive) + 거짓 양성 (False Positive)}}$$

재현율(Recall)은 실제 양성인 샘플 중에서 모델이 양성으로 올바르게 예측한 샘플의 비율을 의미합니다. 의료 영상 분석에서는 재현율이 중요한데, 이는 폐렴 환자를 놓치지 않고 진단하는 것이 중요한 목표이기 때문입니다.

$$\text{재현율} = \frac{\text{참 양성 (True Positive)}}{\text{참 양성 (True Positive) + 거짓 음성 (False Negative)}}$$

F1-score는 정밀도와 재현율의 조화 평균으로, 두 지표 사이의 균형을 평가하는 데 사용됩니다. F1-score는 특히 정밀도와 재현율 사이의 균형이 필요할 때 유용합니다.

$$F1\text{-score} = 2 * \frac{\text{정밀도} * \text{재현율}}{\text{정밀도} + \text{재현율}}$$

F1-score는 본 연구에서 모델이 폐렴을 진단할 때 얼마나 효율적으로 균형 잡힌 성능을 내는지를 평가하는 데 중요한 지표로 사용되었습니다.

4.3. AUC-ROC Curve

AUC-ROC 곡선은 모델의 분류 성능을 평가하는 중요한 지표로, ROC 곡선(Receiver Operating Characteristic curve)은 참 양성 비율(True Positive Rate, TPR)과 거짓 양성 비율(False Positive Rate, FPR)의 관계를 그래프로 나타냅니다. AUC(Area Under Curve)는 ROC 곡선 아래의 면적을 의미하며, 이 값이 1에 가까울수록 모델의 분류 성능이 뛰어납니다.

- TPR(참 양성 비율, 재현율): 실제 양성 샘플 중에서 올바르게 분류된 비율
- FPR(거짓 양성 비율): 실제 음성 샘플 중에서 잘못 분류된 비율

AUC-ROC 값이 1에 가까울수록 모델의 분류 성능이 뛰어나다고 평가됩니다. 본 연구에서는 AUC-ROC를 통해 모델의 전반적인 성능을 시각적으로 평가하였으며, 각 모델의 분류 능력을 종합적으로 분석하는데 사용되었습니다.

이러한 성능 평가 지표들을 통해 본 연구에서는 모델의 분류 성능을 종합적으로 분석하였으며, 각 모델이 폐렴과 정상 상태를 얼마나 정확히 분류하는지 평가할 수 있었습니다. 각 지표는 의료 영상 분석에서 중요한 역할을 하며, 특히 폐렴과 같은 중요한 질병을 진단하는 데 있어 모델의 실질적인 성능을 확인하는 데 필수적입니다.

IV. 실험 및 결과

1. 1차 실험: 모델 성능 비교

본 연구의 1차 실험에서는 CheXNet, ResNet-50, EfficientNetB0 세 가지 CNN 모델을 비교하여 폐렴 진단 성능을 평가했습니다. 각 모델은 동일한 데이터셋에서 학습하였으며, 학습 과정 중 모델의 성능을 학습 데이터와 검증 데이터에 대해 분석하였습니다. 이를 통해 각 모델의 분류 성능을 비교하고, 최종적으로 어떤 모델이 가장 우수한 성능을 보이는지 평가하였습니다.

1.1. 학습 데이터와 검증 데이터 분리

학습을 위해 전체 데이터셋을 학습 데이터, 검증 데이터, 테스트 데이터로 분리하였습니다. 학습 데이터는 모델을 훈련시키기 위해 사용되었으며, 검증 데이터는 학습 중 모델의 성능을 평가하고, 과적합을 방지하기 위해 사용되었습니다. 테스트 데이터는 학습이 완료된 후 모델의 최종 성능을 평가하는 데 사용되었습니다. 데이터 분포는 다음과 같습니다:

- Train Set: 4,215장 (정상: 1,341장, 폐렴: 2,874장)
- Validation Set: 585장 (정상: 8장, 폐렴: 8장)
- Test Set: 1,056장 (정상: 234장, 폐렴: 390장)

데이터의 클래스 불균형 문제를 해결하기 위해, 데이터 증강 기법을 사용하여 학습 데이터의 다양성을 증가시켰으며, 클래스 가중치를 적용하여 모델이 특정 클래스에 치우치지 않도록 하였습니다.

1.2. CheXNet, ResNet-50, EfficientNetB0 성능 비교

본 연구에서는 흉부 X-ray 이미지 기반 폐렴 진단 성능을 평가하기 위해 CheXNet, ResNet-50, EfficientNetB0 세 가지 CNN 모델을 비교하였습니다. 각 모델은 동일한 데이터셋을 사용해 학습되었으며, 평가 지표로는 정확도(Accuracy), 정밀도(Precision), 재현율(Recall), F1-score를 사용하였습니다. 특히 폐렴 클래스와 정상 클래스에 대한 성능을 각각 따로 평가하여 각 모델이 두 클래스에서 어떻게 작동하는지 명확히 비교했습니다.

폐렴 클래스에 대한 평가 결과는 아래의 표에서 확인할 수 있습니다:

모델	정확도	정밀도	재현율	F1-score
CheXNet	0.89	0.93	0.89	0.91
ResNet-50	0.77	0.79	0.85	0.82
EfficientNetB0	0.62	0.62	1.0	0.77

<표3 - 폐렴 클래스에 대한 결과 및 정확도>

- CheXNet은 폐렴 클래스에서 가장 높은 정확도(89%), 정밀도(93%), F1-score(91%)를 기록했습니다. 이는 CheXNet이 폐렴 환자를 매우 높은 신뢰도로 예측할 수 있음을 보여줍니다.
- ResNet-50은 폐렴 클래스에서 77%의 정확도를 보였으며, F1-score도 82%로 CheXNet에 비해 성능이 낮았습니다.
- EfficientNetB0는 재현율(Recall) 측면에서는 100%로, 모든 폐렴 환자를 올바르게 예측하였지만, 정밀도가 62%로 낮아 과다 진단(즉, 정상인도 폐렴으로 분류하는 오류)을 많이 범했습니다.

정상 클래스에 대한 성능 비교는 아래의 표와 같습니다:

모델	정밀도	재현율	F1-score
CheXNet	0.83	0.89	0.86
ResNet-50	0.72	0.62	0.67
EfficientNetB0	0.0	0.0	0.0

<표4 - 정상 클래스에 대한 결과>

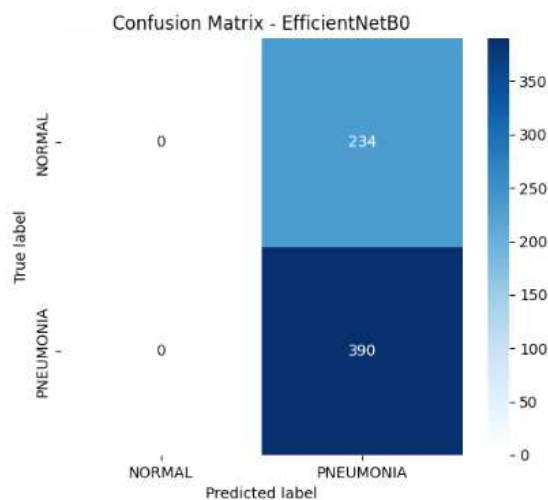
- CheXNet은 정상 클래스에서도 높은 성능을 기록했습니다. 정밀도는 83%, 재현율은 89%로, CheXNet은 정상인도 비교적 정확하게 분류할 수 있었습니다.
- ResNet-50은 정상 클래스에서의 성능이 다소 떨어지며, 재현율이 62%로 정상 환자 중 약 38%를 폐렴으로 오분류했습니다.
- EfficientNetB0는 정상 클래스에서 전혀 올바르게 예측하지 못해 정밀도, 재현율, F1-score 모두 0을 기록했습니다. 이는 EfficientNetB0가 정상 상태를 예측하는 데 심각한 한계를 보였다는 것을 의미합니다.

세 모델을 종합적으로 비교한 결과, CheXNet이 폐렴과 정상 클래스 모두에서 가장 균형 잡힌 성능을 보였습니다. 특히, CheXNet은 두 클래스에서 모두 높은 정밀도와 재현율을 보여주며, 의료 영상 분석에 적합한 모델임을 확인할 수 있습니다. ResNet-50은 비교적 높은 성능을 보였지만, 정상 클래스에서의 성능이 부족한 부분이 있습니다. EfficientNetB0는 폐렴 클래스를 감지하는 데는 뛰어났으나, 정상 클래스를 예측하는 데 큰 어려움을 겪었습니다.

1.3. Confusion Matrix 결과

Confusion Matrix는 각 모델이 실제 정상(Normal)과 폐렴(Pneumonia)을 얼마나 정확하게 분류했는지 시각적으로 보여줍니다. 이를 통해 각 모델의 분류 오류 유형을 확인할 수 있습니다. 아래의 Confusion Matrix 그래프는 각 모델이 정상과 폐렴 환자를 어떻게 예측했는지를 보여줍니다.

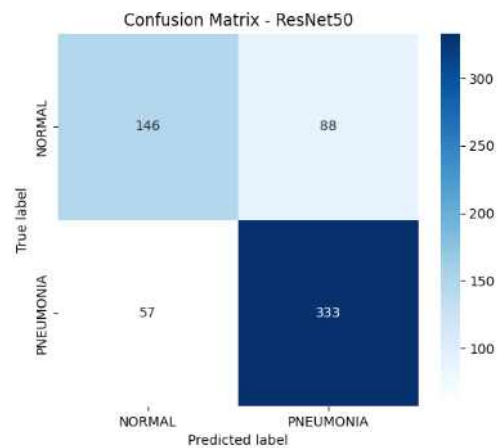
EfficientNetB0 Confusion Matrix



<그림5 - EfficientNetB0 Confusion Matrix>

EfficientNetB0는 정상 환자를 전혀 맞추지 못하고 모두 폐렴으로 예측했으며, 폐렴 환자는 모두 정확하게 예측했습니다. 이는 정상 클래스를 잘못 예측한 비율이 매우 높음을 의미합니다.

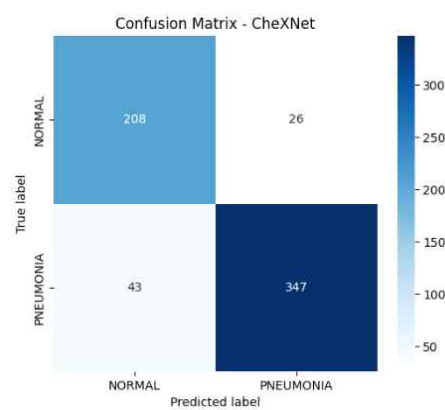
ResNet-50 Confusion Matrix



<그림6 - ResNet-50 Confusion Matrix>

ResNet-50는 정상 클래스에서 146명을 정확히 예측했지만, 88명의 정상 환자를 폐렴으로 오분류했습니다. 폐렴 클래스에서는 333명을 올바르게 예측했고, 57명을 놓쳤습니다.

CheXNet Confusion Matrix



<그림7 - CheXNet Confusion Matrix>

CheXNet은 전체적으로 가장 우수한 성능을 보였습니다. 정상 클래스에서는 208명을 정확히 예측했고, 폐렴 클래스에서도 347명을 올바르게 예측했습니다.

1.4. 손실 및 정확도 그래프

모델의 학습 과정 중 손실(loss) 값과 정확도(accuracy)의 변화를 분석하여, 각 모델이 얼마나 효과적으로 학습했는지를 평가하였습니다. 아래는 각 모델의 손실 및 정확도 그래프입니다.

정확도 그래프 (Accuracy Comparison)

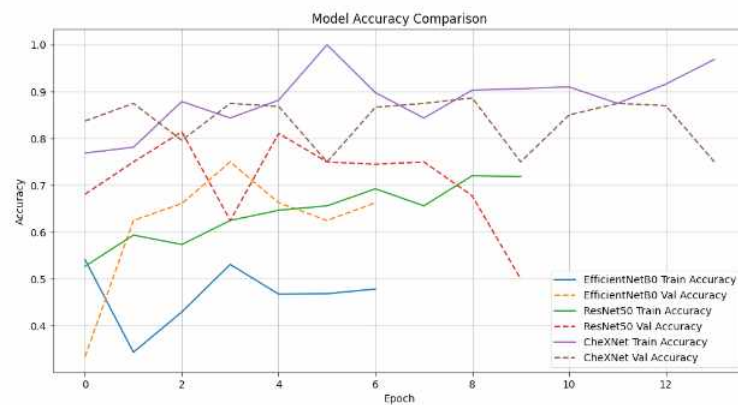
- CheXNet은 학습과 검증 과정에서 빠르게 정확도가 상승하며, 최종적으로 높은 성능에 도달했습니다. 검증 정확도와 학습 정확도 간의 차이가 작아 과적합 없이 학습이 잘 이루어졌습니다.
- ResNet-50은 검증 정확도와 학습 정확도 간에 차이가 있었으며, 검증 정확도가 불안정한 모습을 보였습니다. 이는 과적합 가능성을 시사합니다.
- EfficientNetB0는 학습 정확도는 상승했지만, 검증 정확도는 낮은 수준에 머물렀습니다. 특히 정상 클래스를 예측하지 못한 것이 전체 성능에 큰 영향을 미쳤습니다.

손실 그래프 (Loss Comparison)

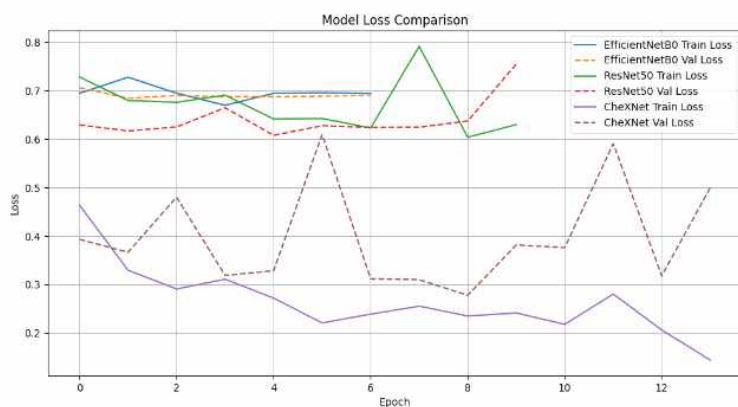
- CheXNet의 손실 값은 학습과 검증 모두에서 꾸준히 감소하였으며, 검증 손실이 낮은 수준에서 안정적으로 유지되었습니다.
- ResNet-50은 검증 손실이 다소 불안정하게 변화하는 모습을 보였으며,

학습 손실과의 차이가 컸습니다. 이는 일부 과적합을 의미할 수 있습니다.

- EfficientNetB0는 학습 손실이 크게 감소했지만, 검증 손실이 매우 높은 수준에서 유지되어, 학습 과정 중 정상 클래스에 대한 예측 오류가 지속되었음을 시사합니다.



<그림8 - 정확도 그래프>



<그림9 - 손실 그래프>

2. 2차 실험: CheXNet 모델 하이퍼파라미터 튜닝

2차 실험에서는 CheXNet 모델의 하이퍼파라미터를 조정하여 성능을 최적화하고, 흉부 X-ray 이미지를 기반으로 폐렴을 진단하는 정확도를 높이는 것을 목표로 하였습니다. 특히, 학습률, 배치 사이즈, 에포크 수 등의 하이퍼파라미터를 조정하여 성능 향상을 도모하였으며, 최종적으로 Confusion Matrix를 통해 성능을 평가하였습니다.

2.1. 하이퍼파라미터 조정: 학습률, 배치 사이즈, 에포크 수

학습률(Learning Rate): CheXNet 모델의 학습률은 $1e-4$ 로 설정하여 비교적 느린 학습 속도로 안정적인 성능 향상을 도모하였습니다. 학습률이 너무 크면 학습이 불안정해질 수 있으므로, 낮은 학습률을 통해 천천히 최적의 가중치로 수렴하도록 설정하였습니다.

배치 사이즈(Batch Size): 배치 사이즈는 32로 설정하였으며, 이는 메모리와 연산 속도의 균형을 고려한 선택이었습니다. 배치 크기가 너무 크면 학습 속도는 빨라질 수 있지만, 모델이 수렴하는 데 어려움이 있을 수 있습니다. 적당한 배치 사이즈는 일반화 성능 향상에 도움이 됩니다.

에포크 수(Epochs): 총 20 에포크 동안 학습을 진행하였습니다. 실험에서는 학습이 충분히 진행될 수 있도록 에포크 수를 설정하였고, 학습 후반부에서 모델의 성능이 더 이상 개선되지 않는 시점을 확인했습니다. 학습 후반에 학습률 스케줄러를 적용하여 학습률을 7 에포크마다 감소시키는 전략을 사용했습니다.

2.2. 조정된 CheXNet 모델 성능 분석

학습 및 검증 과정에서의 손실(Loss)과 정확도(Accuracy)의 변화를 그래프로 분석하여 조정된 하이퍼파라미터가 CheXNet의 성능에 미치는 영향을 평가했습니다.

손실 및 정확도 변화 분석

Train and Validation Loss:

학습 손실은 초기에 급격히 감소하였고, 학습 후반부에 비교적 안정적으로 수렴하였습니다. 검증 손실도 에포크가 진행됨에 따라 감소하여 모델의 성능이 지속적으로 향상되었음을 확인할 수 있었습니다.

Train and Validation Accuracy:

학습 정확도는 빠르게 상승하였으며, 검증 정확도 또한 유사한 패턴으로 증가하였습니다. 최종적으로 CheXNet은 매우 높은 검증 정확도에 도달하여, 정상 및 폐렴 환자 모두를 효과적으로 예측할 수 있음을 확인했습니다.

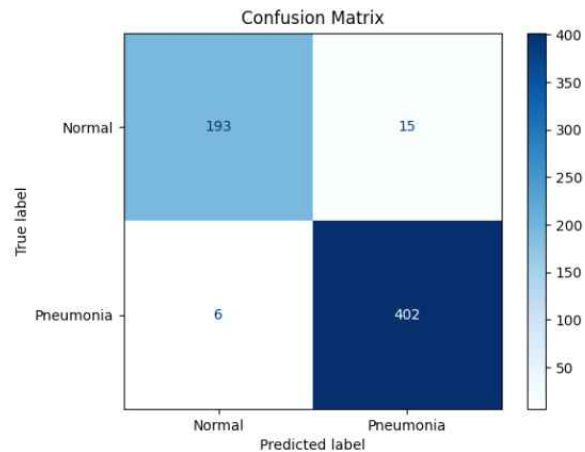
학습과 검증 과정에서 손실과 정확도 그래프를 통해 모델의 성능을 시각적으로 평가할 수 있으며, 학습이 안정적으로 이루어졌음을 확인할 수 있었습니다.

2.3. Confusion Matrix 및 성능 평가

2차 실험에서는 CheXNet 모델을 기반으로 혼동 행렬(Confusion Matrix)과 성능 지표를 분석하여 정상(Normal)과 폐렴(Pneumonia) 클래스를 얼마나 정확하게 분류했는지를 평가했습니다. 혼동 행렬을 통해

모델의 예측 성능을 시각적으로 확인할 수 있으며, 이를 바탕으로 정밀도(Precision), 재현율(Recall), F1-Score 등의 성능 지표를 평가하였습니다.

Confusion Matrix 분석



<그림10 - Confusion Matrix>

정상(Normal): CheXNet 모델은 193명의 정상 환자를 정확하게 예측하였고, 15명의 정상 환자를 폐렴으로 오분류하였습니다.

폐렴(Pneumonia): 402명의 폐렴 환자를 정확하게 예측하였으며, 6명의 폐렴 환자를 정상으로 잘못 분류하였습니다.

이 결과는 CheXNet이 정상 환자와 폐렴 환자를 매우 높은 정확도로 예측할 수 있음을 보여줍니다. 특히 폐렴 클래스에서의 재현율이 매우 높아 폐렴 환자를 놓치지 않고 예측하는 성능이 우수하다는 것을 확인할 수 있습니다.

성능 지표 평가

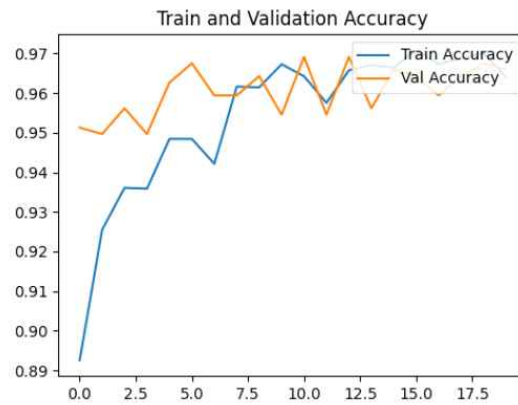
정확도(Accuracy): CheXNet 모델은 전체적으로 97%의 정확도를 기록하였습니다. 이는 데이터셋의 정상 및 폐렴 클래스를 매우 효과적으로 분류하는 모델임을 의미합니다.

정밀도(Precision): 정상 환자를 진단할 때 92.78%의 정밀도를 기록하였으며, 폐렴 환자를 진단할 때 96.39%의 높은 정밀도를 보였습니다. 이는 폐렴 진단 시 과다 진단(정상 환자를 폐렴으로 잘못 진단하는 경우)이 매우 적음을 의미합니다.

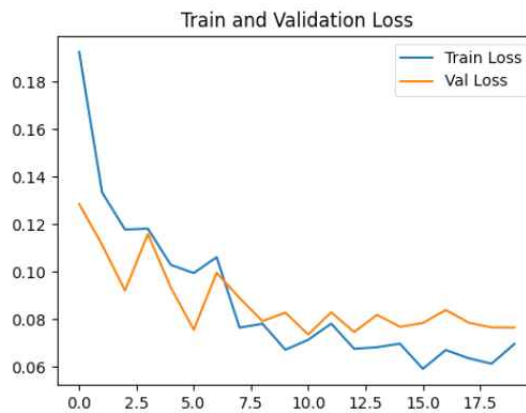
재현율(Recall): 폐렴 환자의 재현율은 98.53%로, 거의 모든 폐렴 환자를 올바르게 진단하였습니다. 이는 CheXNet이 폐렴 환자를 놓치지 않고 진단할 수 있는 매우 중요한 성능 지표입니다.

F1-Score: F1-Score는 정밀도와 재현율의 조화 평균으로, 모델의 전반적인 성능을 평가할 수 있는 지표입니다. 폐렴 환자의 F1-Score는 97.44%로, 매우 높은 성능을 보였습니다.

손실 및 정확도 분석



<그림11 - 정확도 그래프>



<그림12 - 손실 그래프>

Train and Validation Loss: 학습 및 검증 손실은 꾸준히 감소하며 안정적인 수렴 과정을 보여주었습니다. 학습 초기에는 손실이 급격히 감소했으며, 이후 검증 손실도 학습과 유사한 패턴으로 감소하여 모델이 잘 학습되었음을 확인할 수 있었습니다.

Train and Validation Accuracy: CheXNet 모델은 학습 및 검증 정확도 모두에서 96% 이상의 높은 성능을 기록하였으며, 학습이 진행될수록 빠르게 정확도가 상승하는 모습을 보였습니다. 학습 후반부에는 학습과 검증의 정확도가 거의 동일하게 수렴하여 과적합 없이 안정적으로 학습된 것을 확인할 수 있습니다.

3. 최종 결과 분석 및 비교

CheXNet 모델을 사용한 실험에서, 하이퍼파라미터 조정 전후 성능 변화와 다른 모델들과의 비교를 통해 최종 성능을 평가하였습니다. 이를 통해 모델의 성능을 종합적으로 분석하고, 어떤 하이퍼파라미터 조정이 최적의 성능을 이끌어냈는지 확인할 수 있었습니다.

3.1. CheXNet 모델의 최종 성능

2차 실험에서 최종 조정된 CheXNet 모델은 정상 및 폐렴 환자를 분류하는 데 있어 매우 높은 성능을 발휘했습니다. 다음은 CheXNet 모델의 최종 성능입니다:

정확도(Accuracy): 97%

CheXNet 모델은 전체 테스트 데이터셋에서 97%의 정확도를 기록하여, 대부분의 환자를 올바르게 분류하였습니다.

정밀도(Precision): 정상 환자는 92.78%, 폐렴 환자는 96.39%로 높은 정밀도를 기록했습니다.

이는 CheXNet 모델이 정상 환자와 폐렴 환자를 정확하게 구분하는 데 뛰어난 성능을 보였음을 나타냅니다.

재현율(Recall): 폐렴 환자의 재현율은 98.53%로 매우 높은 성능을 보였습니다.

CheXNet 모델은 거의 모든 폐렴 환자를 진단할 수 있었습니다.

F1-Score: 97.44%의 F1-Score로, 정밀도와 재현율 간의 균형 잡힌 성능을 확인할 수 있었습니다.

CheXNet 모델은 일반화 성능이 뛰어나며, 폐렴 환자를 놓치지 않고 정확하게 진단하는 데 탁월한 성능을 보여주었습니다.

3.2. 하이퍼파라미터 조정 전후 성능 비교

하이퍼파라미터 조정 전과 후의 성능을 비교함으로써, 모델의 성능이 어떻게 개선되었는지 분석할 수 있습니다. 주요 하이퍼파라미터로는 학습률, 배치 사이즈, 그리고 클래스 가중치가 사용되었습니다.

Metric	하이퍼파라미터 조정 전	하이퍼파라미터 조정 후
정확도(Accuracy)	89%	97%
정밀도(Precision) - 정상	83%	92.87%
정밀도(Precision) - 폐렴	93%	96.39%
재현율(Recall) - 정상	89%	89%
재현율(Recall) - 폐렴	89%	98.53%
F1-Score - 정상	86%	86%
F1-Score - 폐렴	91%	97.44%

<표5 - 하이퍼파라미터 조정 전후 성능 비교표>

하이퍼파라미터 조정 후, 특히 폐렴 클래스에서 재현율(Recall)이 크게 향상되었습니다. 이는 클래스 불균형 문제를 해결하기 위해 폐렴 클래스에 더 높은 가중치를 부여한 것이 큰 효과를 발휘했음을 나타냅니다.

결과적으로 모델의 전반적인 성능이 크게 개선되었습니다.

3.3. 모델별 성능 평가 요약

CheXNet 모델의 성능을 ResNet-50 및 EfficientNetB0 모델과 비교하여, 각각의 모델이 어떻게 다른 성능을 보였는지 요약할 수 있습니다.

Metric	CheXNet	RestNet-50	EfficientNetB0
정확도(Accuracy)	97%	77%	62%
정밀도(Precision) - 정상	92.78%	72%	0%
정밀도(Precision) - 폐렴	96.39%	79%	62%
재현율(Recall) - 정상	89%	62%	0%
재현율(Recall) - 폐렴	98.53%	85%	100%
F1-Score - 정상	97.44%	67%	0%
F1-Score - 폐렴	97.44%	82%	77%

<표6 - 모델별 성능 평가 요약>

CheXNet 모델이 세 모델 중 가장 뛰어난 성능을 발휘하였으며, 특히 폐렴 진단에서 탁월한 성능을 기록했습니다. ResNet-50은 비교적 높은 성능을 보였지만, CheXNet에 비해 성능이 떨어졌으며, EfficientNetB0는 폐렴 환자를 모두 예측했으나 정상 환자를 잘못 분류하는 큰 오류를 보였습니다.

결론적으로, CheXNet 모델은 정상과 폐렴 환자를 정확하게 예측할 수 있는 성능을 지니고 있으며, 조정된 하이퍼파라미터를 통해 가장 높은 성능을 달성하였습니다.

V. 결론

1. 결론

본 연구에서는 흉부 X-ray 이미지를 기반으로 폐렴을 진단하기 위해 CheXNet, ResNet-50, EfficientNetB0 세 가지 CNN 모델을 비교 분석하였습니다. 실험 결과, CheXNet 모델이 가장 우수한 성능을 보였으며, 정상과 폐렴 환자를 높은 정확도로 분류할 수 있음을 확인하였습니다.

CheXNet 모델은 특히 폐렴 환자를 놓치지 않고 예측할 수 있는 높은 재현율(98.53%)을 기록하였으며, 전체 정확도는 97%에 달했습니다. 또한 정밀도와 F1-score에서도 가장 우수한 성능을 보였습니다.

ResNet-50과 EfficientNetB0 모델은 상대적으로 낮은 성능을 보였으며, 특히 EfficientNetB0는 정상 환자를 분류하는 데 많은 오류를 발생시켰습니다.

본 연구의 결과로, CheXNet 모델이 흉부 X-ray 이미지를 기반으로 한 폐렴 진단 시스템에서 가장 적합한 모델임을 알 수 있었습니다. 특히, 폐렴 환자를 빠르고 정확하게 진단할 수 있어 의료 영상 분석 분야에서 활용 가능성이 큼니다.

2. 한계점 및 발전방향

한계점

데이터셋의 제한: 본 연구에서는 특정 흉부 X-ray 데이터셋을 사용하였으며, 이 데이터셋의 크기와 다양성에 따라 모델의 성능이 결정되었습니다. 데이터셋이 충분히 크지 않거나, 다양한 임상 데이터를 포함하지 못할 경우, 모델의 일반화 성능이 저하될 수 있습니다.

모델의 해석 가능성 부족: CNN 모델은 일반적으로 블랙박스 모델로, 의료 전문가들이 진단 과정에서 모델의 결정을 이해하고 신뢰하기 어렵습니다. 이러한 해석 가능성의 부족은 실제 임상에서 사용될 때 한계로 작용할 수 있습니다.

발전방향

데이터셋 확장: 향후 연구에서는 다양한 병원 및 환자 데이터를 포함하는 대규모 데이터셋을 사용하여 모델을 학습시키는 것이 필요합니다. 특히, 다양한 연령대, 성별, 인종을 포함한 데이터를 추가하면 모델의 일반화 성능을 높일 수 있습니다.

모델의 해석 가능성 향상: 모델의 결정을 설명할 수 있는 XAI(Explainable AI) 기술을 도입하여, 의료 전문가들이 모델이 특정 폐렴 패턴을 어떻게 인식했는지 시각적으로 이해할 수 있도록 하는 것이 필요합니다. 이를 통해 모델에 대한 신뢰성을 높일 수 있을 것입니다.

실시간 진단 시스템 개발: 본 연구의 결과를 바탕으로, 실시간으로 X-ray 이미지를 분석하고 폐렴 진단을 도울 수 있는 임상 도구를 개발

할 수 있습니다. 모바일 기기나 병원 시스템에 적용할 수 있는 경량화된 버전의 모델을 개발하는 것이 필요합니다.

다중 질환 진단 모델: CheXNet은 폐렴뿐만 아니라 다른 흉부 질환도 진단할 수 있는 확장성이 있습니다. 향후 연구에서는 다중 질환을 진단할 수 있는 모델로 확장하여, 종합적인 흉부 질환 진단 시스템을 구축할 수 있습니다.

참고 문헌

- [1] He, K., Zhang, X., Ren, S., and Sun, J. "Deep Residual Learning for Image Recognition," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778, 2016.
- [2] Huang, G., Liu, Z., van der Maaten, L., and Weinberger, K. Q. "Densely Connected Convolutional Networks," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4700–4708, 2017.
- [3] Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., et al. "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning," arXiv preprint arXiv:1711.05225, 2017.
- [4] Tan, M. and Le, Q. V. "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," Proceedings of the 36th International Conference on Machine Learning (ICML), pp. 6105–6114, 2019.
- [5] Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C. S., et al. "Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning," Cell, vol. 172, no. 5, pp. 1122–1131.e9, 2018.
- [6] Wang, X., Peng, Y., Lu, L., Bagheri, M., and Summers, R. M. "ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases," arXiv preprint arXiv:1705.02315, 2017.
- [7] Khan, W., Zaki, N., and Ali, L. "Intelligent Pneumonia Identification from Chest X-Rays: A Systematic Literature Review," IEEE Access, vol. 9, pp. 51747–51771, 2021.
- [8] 최용은, 이승완. "흉부 X선 영상을 이용한 작은 층수 ResNet 기반 폐렴 진단 모델의 성능 평가." 대한방사선과학회지, vol. 46, no. 4, pp. 277–285, 2023.