

영역별 화소값 분석을 통한 흉부 X선 오픈 데이터셋 품질 평가

최현진¹, 배수빈², 전주성³, 이정원^{1,2}

¹아주대학교 전자공학과, ²아주대학교 AI융합네트워크학과,

³아주대학교 의과대학 영상의학과

hjaj12020409@ajou.ac.kr, bpgs5316@ajou.ac.kr, sunnahn@ajou.ac.kr, jungwony@ajou.ac.kr

Quality Evaluation of Chest X-ray Open Dataset through Pixel Value Analysis by Region

Hyeon-Jin Choi¹, Su-Bin Bea², Joo-Sung Sun³, Jung-Won Lee^{1,2}

¹Dept. of Electrical and Computer Engineering, Ajou University

²Dept. of AI Convergence Network, Ajou University

³Dept. of Radiology, Ajou University

요 약

인공지능의 발전으로 의료영상 분야에서 딥러닝 기반 질병 진단 연구가 활발하다. 그러나 모델 개발 시 학습 데이터의 개수와 품질은 매우 중요한데, 의료 분야 특성상 접근 가능한 데이터셋이 적으며 오픈 데이터셋은 서로 다른 기관에서 배포되거나 웹상에서 수집된 것으로 진단에 적합한 품질을 기대하기 어렵다. 또한, 기존 연구는 데이터셋이 학습에 적합한지에 대한 품질검증 없이 사용한다. 따라서 본 논문에서는 임상에서 사용하는 화질 평가 요소에 근거를 두고 영역별 화소값 분석을 통한 흉부 X선 영상 품질 평가 기법을 제안한다. 오픈 데이터셋 JSRT, Chest14와 국내 A 병원 데이터셋 AUH에 제안한 기법을 적용한 결과 민감도 91.5%, 특이도 96.1%의 우수한 성능을 확인하였다.

1. 서론

인공지능 분야는 데이터의 기하급수적인 증가와 함께 지속적인 기술의 발전을 이루며 응용 범위를 넓혀가고 있다[1]. 의료영상 분야에서도 질병 진단에 딥러닝 기술을 적용하려는 연구가 진행되고 있는데, COVID-19 확산의 영향으로 흉부 X선 영상 기반 폐 질환 진단모델의 개발이 활발하다[2].

이미지 분류에 활용되는 지도학습 딥러닝 모델은 주어진 데이터에 대한 반복 예측을 통해 손실(Loss)을 최소화하는 방향으로 파라미터를 갱신한다. 따라서 주어진 데이터에서 스스로 특징을 추출하고 학습하여 판단할 수 있지만, 데이터의 품질이 낮다면 모델의 성능과 신뢰성을 보장할 수 없게 된다.

의료 분야 특성상, 접근 가능한 데이터가 매우 적어 많은 연구에서 오픈 데이터셋을 이용하여 질병 진단모델을 개발한다. 그러나 오픈 데이터셋은 서로 다른 기관에서 배포되거나 웹상에서 수집된 것으로

진단에 적합한 영상 품질을 기대하기 어렵다. 그림 1은 대표적인 흉부 X선 데이터셋 Chest X-ray 14의 예시이다. 그림 1(a)와 같은 고품질 데이터도 존재하지만, 그림 1(b)와 같이 인공 장치(의료기기 등)에 의한 인공음영이 있거나 그림 1(c), (d)와 같이 흉부가 제대로 촬영되지 않은 경우도 존재한다. 인공음영은 진단모델에서 질병으로 오인될 수 있고, 대조도나 자세가 불량한 경우는 모델이 흉부를 제대로 인식할 수 없으므로 학습 데이터로 적절치 않다.



a. 정상 b. 인공음영 c. 대조도 d. 자세 불량
(그림 1) Chest14 품질에 따른 흉부 X선 영상 예

그러나, 이러한 데이터의 품질을 개발자가 결정하기에 의료 분야 도메인 지식에 대한 진입 장벽이 높다. 이로 인해 기존 의료영상 기반 연구는 데이터에 대한 품질검증 없이 필요에 따라 데이터를 전처리하여 사용해왔다. [3]은 JSRT, GDH, SZH 데이터셋에 대해 품질검증 없이 누락 억제 전처리 후 폐 결절 검출 학습 데이터로 사용하였다. 또, [4]는 COVID-19를 진단함에 품질검증을 거치지 않은 데

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터지원사업의 연구결과로 수행되었음 (IITP-2022-2020-0-01461)

이더넷 Cohen JP, Chest X-ray 8을 이용하며 [5]는 급성 호흡곤란 증후군 진단모델 개발을 위해 JSRT, Montgomery 데이터셋에 대한 폐 분할 전처리 알고리즘을 제안하며 품질검증을 진행하지 않는다.

품질를 판단하는 임상 기준 중, 별도의 의학 지식 없이 평가 가능한 세 가지(인공음영, 대조도, 자세)를 노이즈로 분류하고, 영역별 화소값 분석 기반 흉부 X선 영상 품질 평가 기법을 제안한다.

오픈 데이터셋 JSRT, Chest14와 국내 A 병원 데이터셋 AUH에 대한 품질 평가 결과 민감도 91.5%, 특이도 96.1%의 우수한 성능을 확인하였고, 이로부터 오픈 데이터셋 JSRT, Chest14의 품질을 확인하였다. 제안하는 기법은 학습 데이터 품질 평가로 모델의 신뢰도 검증을 돕고, 저품질 영상을 정제하여 모델의 성능 향상을 지원할 수 있다.

2. 흉부 X선 영상 품질 평가 방법

본 장에서는 흉부 X선 영상 품질 평가 방법을 제안하며, 요약하면 그림 2와 같다.

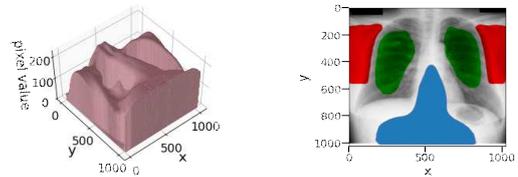
2.1 흉부 X선 영상 품질 평가 요소 분석

국내 A 병원의 흉부 X선 영상 화질 평가 기준을 기반으로 품질 평가 요소를 분석, 설정한다. 품질 평가 요소는 일반 정보와 영상 정보로 나뉘는데, 일반 정보는 환자의 개인정보 등 영상 자체와 연관이 없어 평가 요소에서 제외하였다. 영상 정보는 인공음영, 포함범위, 환자 자세, 흡기 정도, 투과상태로 분류된다. 이 중 포함범위와 흡기 정도는 폐 영역이 얼마만큼 촬영되었는지를 평가하는 항목으로 개발자가 정확한 판별 여부를 검증하기 어렵다. 따라서 본 논문에서는 이를 제외한 인공음영, 환자 자세, 투과상태(대조도)를 평가 요소로 설정하였다.

2.2 흉부 X선 영상 영역분할 기반 품질 평가

그림 3(a)는 전처리한 고품질 흉부 X선 데이터의

좌표별 평균 밝기를 시각화한 분포도이다. z축은 밝기 정도인 화소값을 나타낸다. 영상은 z값에 따라 세 영역으로 구분되며, 그림 3(b)는 분할 한 결과이다(파란색: 심장, 빨간색: 견갑골, 초록색: 폐).

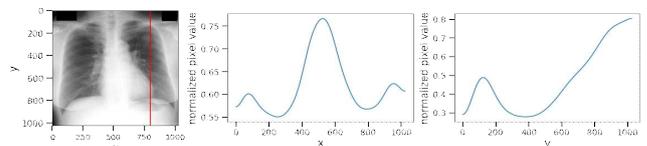


a. 영상 평균 밝기 분포도 b. 밝기에 따른 영역분할 (그림 3) 흉부 X선 영상 평균 밝기에 따른 영역분할

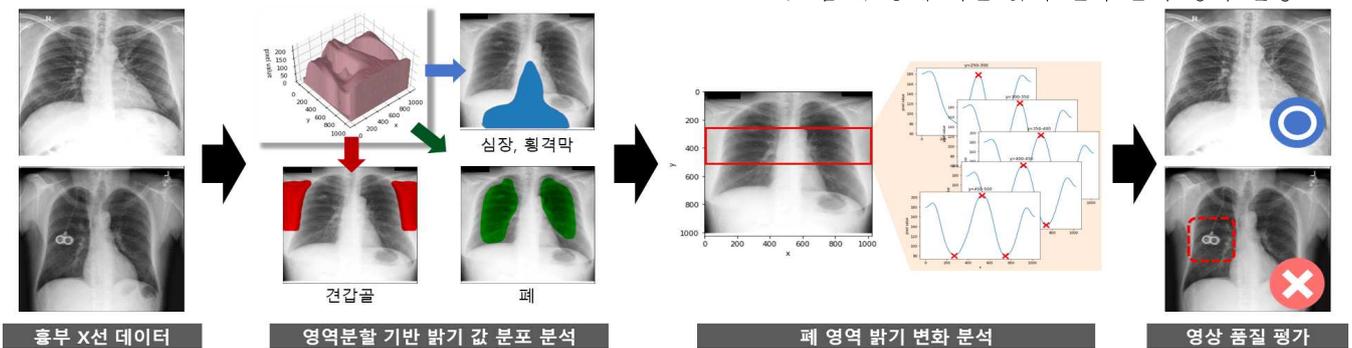
흉부 X선 촬영은 폐·흉곽 질환 진단을 위한 검사이다. 따라서 폐와 폐의 호흡운동을 돕는 횡격막, 흉곽 내 심장은 모든 영상에서 관찰된다. 견갑골 또한 모든 영상에 대해 같은 위치에서 관찰되는 요소로 A 병원 영상 화질 평가 기준에서도 견갑골의 위치로부터 환자 자세를 평가한다. 따라서 그림 3(b) 세 영역을 흉부 X선 영상 내 주요 요소로 결정하고, 각 영역에 대한 품질 기준 정의 및 평가로 영역 내 노이즈 검출, 주요 요소 위치 확인을 수행한다.

2.2 폐 영역 밝기 변화 분석 기반 품질 평가

그림 4는 밝기 변화 분석 영역 설정 과정을 나타낸다. 먼저, 그림 4(b)는 흉부 영상의 각 x 좌표에 대한 모든 y 좌표(0-1023)에서의 화소값 합의 평균을 정규화한 그래프이다. 이를 통해 통계적인 폐 중심을 결정하였고, 그림 4(a)의 빨간 선이 그 예이다. 그다음, 폐 중심에서의 통계적인 밝기 분포를 정규화하여 그래프로 나타내었고 이는 그림 4(c)와 같다. 이때, 쇄골과 심장 근처 영역은 혈관, 늑골에 의해 다른 영역과 밝기 분포가 다르게 나타나므로 분석에



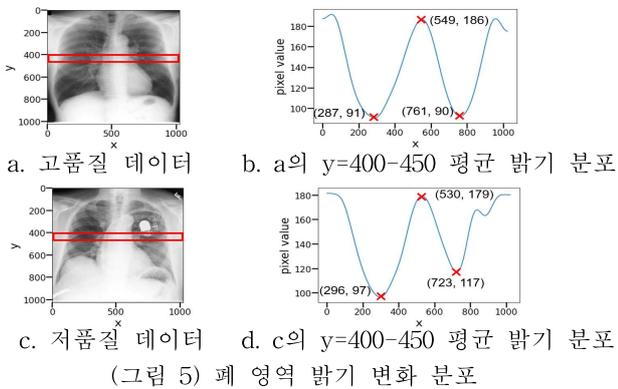
a. 고품질 영상 b. a의 평균 밝기변화 c. 폐 중심 밝기변화 (그림 4) 통계 기반 밝기 변화 분석 영역 설정



(그림 2) 흉부 X선 영상 품질 평가 기준 정의 절차

제외하였다. 따라서 그림 4(c)에서 정규화된 밝기 값이 0.3 이하인 범위를 통계적으로 쇄골, 심장에 방해받지 않는 폐 영역으로 판단, 밝기 변화 분석 영역으로 설정하였다(x: 0-1023, y: 250-500).

그림 5는 설정한 영역 중 y=400-450의 평균 밝기 분포 그래프이다. 그림 5(b)를 보면 몸의 윤곽과 척추 영역에서 극댓값, 폐 영역에서 극솟값을 갖는다. 따라서 몸의 윤곽, 척추, 폐 영역에서의 극댓값, 극솟값 존재 여부와 둘 간의 차를 확인함으로써 인공음영의 유무와 대조 정도를 판별할 수 있다. 그림 5(b), (d)의 극댓값과 극솟값의 차를 보면, 고품질은 95, 96, 저품질은 82, 62로 인공음영이 존재하는 우측 폐에서 낮은 값을 보이는 것을 확인할 수 있다.



(그림 5) 폐 영역 밝기 변화 분포

3. 실험 결과

3.1 데이터셋 사양 분석

오픈 데이터셋 JSRT, Chest14와 국내 A 병원 데이터셋 AUH에 대한 적용 실험으로 검증하였으며, 실험에 사용된 데이터는 표 1과 같다.

<표 1> 데이터 레이블

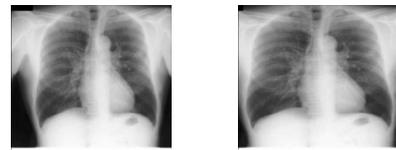
분류	데이터셋	개수	
고품질	JSRT-A	정상	244
	AUH-A	정상	2,514
저품질	Chest14-B	2,000	
		인공음영	1,707
		대조도	893
	AUH-B	환자 자세	75
		140	
		인공음영	2
	환자 자세	138	

3.2 흉부 X선 영상 전처리 및 정규화

품질 평가에 앞서, 영상 전처리 및 정규화를 수행한다. JSRT의 크기는 (1024, 1024, 3), Chest14 크기는 (2048, 2048, 3)이며 AUH는 영상마다 다르다. X선 영상은 회색조이므로 RGB 차원을 제거하고 크기를 조정하여 (1024, 1024)로 재설정하였다.

흉부 X선 영상은 촬영 장비, 환경 등에 따라 선명도와 대비 정도가 달라, 데이터셋마다 다른 화소값 범위를 갖는다. 그러나 각 영상이 다른 범위의 화소값을 가질 경우, 하나의 밝기 분석 기반 품질 평가 기준을 정의하기 어렵다. 따라서 모든 영상이 같은 화소값 범위를 가지며 고품질 데이터의 특성을 유지하도록 각각을 고품질 데이터셋의 평균 화소값 범위(13-230)로 재조정 하였다.

그다음, 흉곽 추출을 수행하였다. 척추 기준 좌우 가장 밝은 지점을 몸의 윤곽으로 설정, 영상 내 배경의 비율을 조정한다. 고품질 데이터셋의 통계적인 쇄골, 횡격막 범위(y=150-737)를 폐 영역으로 보고 중간 지점(y=444)을 기준으로 잡는다.



a. 전처리 및 정규화 전 b. 전처리 및 정규화 후 (그림 6) 흉부 X선 영상 전처리 및 정규화 예시

3.3 흉부 영역분할 기반 품질 평가

영역분할 기반 품질 평가로 주요 요소의 배열 확인 및 영역별 노이즈 검출을 진행한다. 제한한 방법에 따라 영상을 세 영역으로 분할하고, 고품질 데이터에 대한 통계분석을 통해 영역별 품질 평가 기준을 확립하였다. 각 영역에서 평가 기준치를 벗어나는 화소 개수가 오차범위 이상인 데이터를 노이즈로 분류한다. 영역별 품질 평가 기준은 표 2와 같다.

<표 2> 영역별 품질 평가 임계치

	영역 설정 조건	평가 기준	오차범위
심장, 횡격막	$200 \leq m \leq 255$	$n \leq 50$	15%
	$0 \leq y \leq 1000$		
견갑골	$150 \leq m \leq 199$	$n \leq 60$	25%
	$(x \leq 300) \text{ or } (730 \leq x)$		
폐	$100 \leq y \leq 550$	$220 \leq n$	30%
	$0 \leq m \leq 100$		
	$150 \leq x \leq 870$ $90 \leq y \leq 700$		

* m은 고품질 데이터 평균 밝기 값, n은 화소값.

3.4 흉부 밝기 변화 분석 기반 품질 평가 결과

폐 영역에 대한 더 정밀한 노이즈 검출을 위해 밝기 변화 분석을 수행한다. 정상 데이터를 통계 분석한 결과 통계적인 몸의 윤곽 범위는 $x=0-163$, $x=869-1023$ 이고 척추의 범위는 $x=415-640$ 로 나타났다. 폐 범위는 이를 기반으로 설정하였다. 표 3은 밝기 변화 분포를 평가하는 기준을 나타낸다.

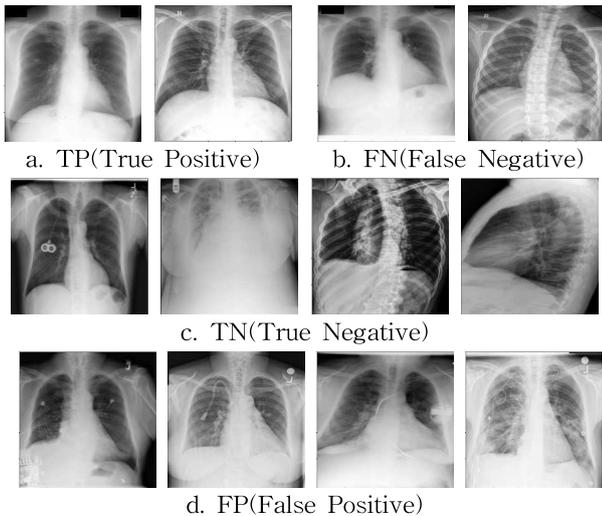
또한, 척추와 폐 영역에서 구한 극댓값, 극솟값의 차는 JSRT-A와 AUH-A의 통계적인 극값 차(111, 135)와 Chest14-B 중 인공음영과 대조도 데이터의 통계적인 극값 차(59) 사이 범위에서 우수한 성능을 보였던 값인 70으로 설정하였다.

<표 3> 밝기 변화 분포 평가 기준

	영역 설정 조건	평가 기준
윤곽	$0 \leq x \leq 163$	극댓값 O
	$869 \leq x \leq 1023$	
척추	$415 \leq x \leq 640$	극댓값 O
폐	$192 \leq x \leq 433$	극솟값 O and 극댓값 X
	$575 \leq x \leq 831$	

3.4 품질 평가 결과

영역별 화소값 분포, 폐 영역 밝기 변화 분석 두 단계에 걸쳐 품질 평가를 진행한 결과 라벨링된 고품질 데이터셋 JSRT-A, AUH-A에 대해 각각 90.9%, 91.6%를 고품질로 옳게 평가하였고, 저품질 데이터셋 Chest14-B, AUH-B에 대해 95.9%, 99.3%를 저품질로 옳게 평가하며 우수한 성능을 보였다. 그림 7은 분류 결과의 예시이다.



(그림 7) 고품질, 저품질 데이터 검출 예시

그림 7(a)는 인공음영이 없고 자세, 대조도가 적절하여 고품질로 평가되었다. 그림 7(b)는 세 요소 모두 해당하지 않으나 저품질로 평가된 경우이다. 좌측 영상은 흡기 문제로 폐 영역이 작게 촬영되어 영역별 평가에서, 우측 영상은 어린아이로 흉부의 형태가 성인과 달라 밝기 변화 분석에서 저품질로 평가되었다. 그림 7(c)는 저품질에 대해 저품질로 옳게 평가한 결과이다. 저품질에 대해 고품질로 잘못 판단된 경우는 인공음영이 매우 작은 경우, 인공음영이 몸의 윤곽에 겹쳐진 경우, 폐가 잘 보이지 않으나 일부 영역에서 매우 어두운 값을 가지는 경우

로 그림 7(d)에서 확인할 수 있다. 자세가 부적절한 경우는 모든 실험 데이터에 대해 옳게 판단되었다.

JSRT-A, Chest14-B, AUH-A, AUH-B에 대한 적용 실험으로 우수한 성능이 검증된 품질 평가 기법을 이용하여 인공음영, 대조도, 환자 자세 측면에서 오픈 데이터셋 JSRT, Chest14에 대한 품질 평가를 진행하였다. 그 결과 JSRT는 전체 247개 데이터 중 90.3%, Chest14는 전체 112,120개 데이터 중 27.0%의 고품질 데이터를 갖는 것으로 평가되었다.

5. 결론

본 논문에서는 흉부 X선 영상에 대한 영역별 화소값 분포, 폐 영역의 밝기 변화 분석을 통한 통계적인 정밀 품질평가 기법을 제안한다. 오픈 데이터셋 JSRT, Chest14와 국내 A 병원의 데이터셋 AUH를 활용하여 적용 실험을 수행한 결과 민감도 91.5%, 특이도 96.1%의 우수한 성능을 확인하였다. 이 기법을 통해 오픈 데이터셋 JSRT, Chest14의 고품질 데이터 비율은 90.3%, 27.0%로 평가되었다. 향후, 추가적인 데이터셋에 대한 실험을 진행하여 기법을 일반화하고, 질병 진단모델의 학습 데이터셋에 대해 본 기법을 적용해 봄으로써 해당 모델의 성능 및 신뢰도 검증에 활용할 예정이다.

참고문헌

[1] Dong, Shi, Ping Wang, and Khushnood Abbas. A survey on deep learning and its applications. Computer Science Review 40, 100379, 2021.

[2] Alyasseri, Zaid Abdi Alkareem, et al. Review on COVID 19 diagnosis models based on machine learning and deep learning approaches. Expert systems 39.3, e12759, 2022.

[3] Li, Xuechen, et al. Multi-resolution convolutional networks for chest X-ray radiograph based lung nodule detection. Artificial intelligence in medicine 103, 101744, 2020.

[4] Ozturk, Tulin, et al. Automated detection of COVID-19 cases using deep neural networks with X-ray images. Computers in biology and medicine 121, 103792, 2020.

[5] Reamaroon, Narathip, et al. Robust segmentation of lung in chest x-ray: applications in analysis of acute respiratory distress syndrome. BMC Medical Imaging 20.1, 1-13, 2020.