

신뢰가능한 골다공증 진단을 위한 Causal 추론 기술 연구

Causal Reasoning for Reliable Osteoporosis Diagnosis

요 약

본 논문에서는 어깨 x-ray 이미지만으로도 골다공증을 정확하게 진단할 수 있는 모델을 개발하고자 한다. 의료 데이터의 특성상, 골다공증 환자의 어깨 x-ray 데이터의 수가 적기 때문에 다양한 시나리오의 자기 지도학습을 통해 적은 데이터로도 높은 성능을 낼 수 있는 모델 학습 방법을 연구한다. Proxy-task와 MAE(Masked Auto Encoder)는 컴퓨터비전 분야에서 모델과 상관없이 그 성능이 보장되는 자기 지도 학습 방법으로, 다양한 분야에서 활용되고 있다. 이에 따라 의료 영상 분야에서도 회전 Proxy-task와 MAE 방식의 사전 학습을 통해 정확도 높은 진단 모델을 개발한다. Accuracy와 ROC-AUC뿐만 아니라 GradCam 기술로 계산한 모델과 의료 전문가의 판단 근거 유사성을 모델 성능 지표로 사용하여, 더욱 신뢰성 있고 정확한 골다공증 진단 모델과 그 학습 방법을 제안하고자 한다.

1. 서 론

골다공증으로 인한 골절은 환자의 생활 질을 크게 저하하며, 심각한 경우 사망에 이르게 할 수도 있다. 이에 골다공증을 초기에 정확하게 진단하는 것의 중요성은 커지고 있다. 전통적으로 골다공증의 진단은 골밀도 측정에 의존해 왔으나, 해당 방법은 비용과 접근성의 문제가 존재한다.

이러한 신속하고 정확한 골다공증 진단의 필요성을 기반으로 본 논문에서는 어깨 X-ray 이미지를 사용하여 골다공증을 진단할 수 있는 모델을 개발하고자 한다. 또한 질병 진단이라는, 판단의 근거가 합리적이고 정확해야 하는 문제의 특성을 고려해 판단의 근거가 causal한 모델의 개발을 목표로 한다.

해당 목표에서 어깨 골다공증을 진단하는 모델을 학습시킬 수 있는 데이터의 수가 절대적으로 적은 것이 가장 중요한 문제점이라고 할 수 있다. 이를 위해 Masked Auto-encoder와 Proxy task의 자기지도학습을 통해 적은 데이터셋에서도 문제를 정확하게 풀기 위한 모델의 학습 방법에 대해 연구하고자 한다. 이와 더불어, 설명 가능한 AI 기술 중 하나인 GradCAM을 활용하여 모델이 이미지의 어느 부분을 보고 결정을 내렸는지 시각적으로 나타내어, 모델의 판단 근거를 투명하게 제시하고자 한다. 이는 의료 전문가들이 AI의 결정을 더 잘 이해하고 신뢰할 수 있게 함으로써, 진단의 정확성을 높이는 데 기여할 수 있다. 해당 연구를 통해 기존의 골다공증 진단 방

법을 효과적으로 보완하여 초기 진단과 치료에 중요한 기여를 할 수 있다. 또한 데이터 효율적인 모델 학습 방법과 해당 방법들의 정확도와 설명 가능성의 비교를 통해 의료분야를 포함한 다양한 문제에서 활용될 수 있다.

2. 연구 배경 및 관련 연구

자기 지도 학습. 자기 지도학습은 라벨이 없는 이미지 데이터셋에서의 모델 학습 방법이다. 이는 데이터 수가 적은 라벨 데이터셋에서 학습하기 전 미리 모델이 타겟 대상에 대한 특징을 사전 학습하도록 한다.

MAE(Masked Auto Encoder)는 이미지의 일부를 마스킹하고 남은 정보를 기반으로 이미지를 재생성하는 과정에서 모델이 중요한 특징을 학습하도록 하는 자기지도학습의 방법이다. 이러한 설계를 통해 훈련 속도를 3배 이상 향상시키고 정확도를 개선할 수 있다. 또한 높은 용량의 모델을 효율적으로 학습할 수 있도록 한다[1].

회전 Proxy-task는 입력 이미지를 회전시킨 후, 모델이 이미지가 회전된 각도를 예측하며 이미지 특징을 파악하도록 하는 자기 지도학습 방법이다. 모델은 회전에 따른 이미지 변형을 인식하고, 이러한 변형이 골다공증의 진단과 어떠한 연관성을 가지는지 학습할 수 있다. 추가적으로, 회전 Proxy-task를 통해 향상된 특징 추출 능력은 모델이 실제 진단 시 골다공증을 보다 정확히 분류하는 데 기여할 것으로 기대된다.

Causal 추론의 의미와 분석. 본 연구에서 Causal 추론이란 모델이 골다공증을 진단하는 판단 근거와 의사의 판단 근거가 유사함을 의미한다. Grad-CAM(Gradient-weighted Class Activation Mapping)은 모델이 이미지의 라벨을 판단할 때 중요한 역할을 한 이미지 영역을 그래디언트 연산을 기반으로 시각화하는 도구이다. 의료 전문가가 골다공증을 진단할 때 근거가 되는 이미지 영역과 Grad-CAM을 통한 모델의 판단 근거 영역이 겹치는 정도를 계산하여, 모델의 Causal 추론 능력을 평가할 수 있다.



그림 1. 어깨 X-ray(왼)와 Grad-CAM을 활용한 골다공증 진단에서의 모델의 판단 근거 영역(오)

3. 연구 방법

학습 데이터. 골다공증 분류 모델 학습을 위한 어깨 X-ray 이미지는 총 1,798장 (골다공증 환자 1020명, 정상 778명)이며, training, validation, test는 0.77, 0.11, 0.12의 비율로 나누었다. 사전 학습에 사용된 흉부 X-ray 이미지는 총 50만 장으로, 공용 데이터 셋인 NIH ChestX-ray14 (75,312장), Stanford CheXpert (191,028장), MIMIC-CXR (243,334장)을 사용하였다. 어깨 X-ray 이미지를 활용한 사전 학습에서는 골다공증 진단 시 활용할 training 이미지 1,384장을 사용하였다.

진단 모델. 골다공증 진단 모델로서 Densenet121과 ViT-small을 사용하였다. 어깨 X-ray의 데이터 수가 부족한 점을 감안하여 흉부 X-ray 이미지에서의 MAE를 통해 사전학습된 두 모델[2]을 기본적으로 사용하였다. 사전학습 이미지를 ImageNet과 같은 일반적인 이미지가 아닌 흉부 X-ray를 사용함으로써, 모델이 X-ray 이미지에서의 특징을 추출하는 데에 더 집중할 수 있다[3].

사전학습 시나리오. 흉부 X-ray 이미지에서 MAE를 통해 사전 학습된 모델을 기본으로 한다. 이에 어깨 X-ray 이미지를 활용한 3가지의 자기 지도학습을 한 번 더 진행한다.

1. 흉부 X-ray MAE
2. 흉부 X-ray MAE + 어깨 X-ray Rotation Proxy
3. 흉부 X-ray MAE + 어깨 X-ray MAE(random crop)
4. 흉부 X-ray MAE + 어깨 X-ray MAE(center crop)

두 단계의 자기 지도 학습을 통해 모델이 골다공증 진단을 위한 특징을 더욱 정교하게 학습할 수 있을 것이라 기대하고 정확도와 Causal 추론 성능을 비교·분석한다. 또한 세 가지의 자기진단학습 시나리오에 따른 성능을 비교하여 정확하고 causal 추론이 가능한 자기지도학습 방법을 파악한다.

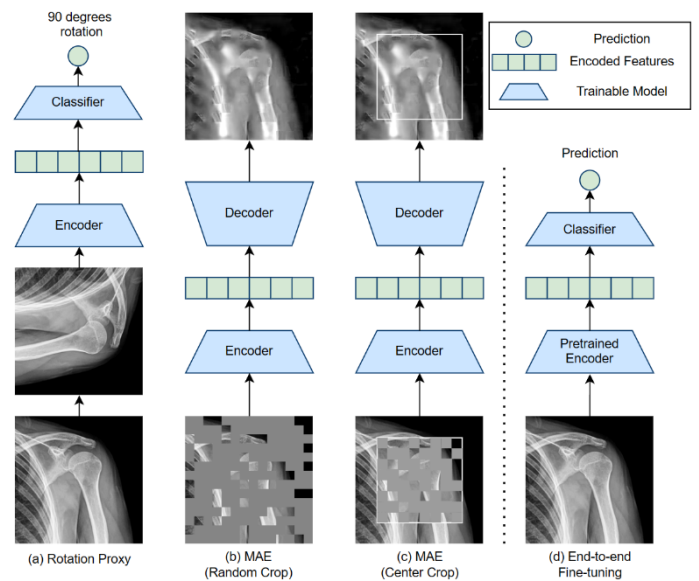


그림 2. 흉부 X-ray MAE 이후의 3가지 사전학습(2,3,4 시나리오)과 미세 조정 학습

MAE & Proxy task. encoder에는 vanilla ViT를 사용하였으며, 어깨 X-ray 이미지의 전처리와 MAE는 흉부 X-ray MAE의 방식과 동일하게 진행하였다. X-ray 이미지가 인접 픽셀과의 정보 유사성이 일반 이미지에 비해 높은 점을 감안하여 masking 비율은 90%로 설정하였다. Masking 비율이 매우 높기 때문에 decoder가 이미지를 재생성하는 성능은 떨어지지만, 이미지의 특성을 추출하는 데에는 더 높은 성능을 보인다 [3] 흉부 X-ray MAE를 진행한 선행 연구에서는 이미지의 모든 부분에서 랜덤하게 resizedcrop과 masking을 진행하였다. 하지만 골다공증 진단 시, 어깨 X-ray 이미지의 중심 영역만이 사용되어 이미지 외곽을 crop 하는 것이 무의미하다. 따라서 전체 이미지의 80%에 해당하는 중심 영역을 먼저 자른 뒤, 이미지에서 0.5~1.0 비율로 랜덤하게 resizedcrop과 masking을 진행하였다. 이는 골다공증을 판단하는 데에 중요한 영역을 집중적으로 masking 할 수 있다. 실험에서는 center crop를 진행한 경우 모델이 중요한 특징을 더욱 잘 추출할 수 있을 것이라 기대하고 random crop한 경우와 성능을 비교·분석하였다. Rotation self-supervision의 경우, 어깨 X-ray 이미지에 대해 0도, 90도, 180도, 270도의 네 가지 회전을 적용하고, 모델이 원본 이미지에서의 회전 각도를 정확히 예측할 수 있도록 학습을 진행하였다.

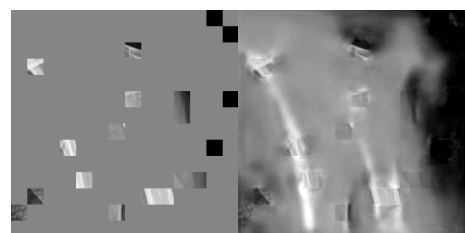


그림 3. Random crop + 90% random masking 이미지(왼)와 모델이 재구성한 어깨 X-ray 이미지(오)

미세조정 학습. 자기 지도학습을 통한 사전학습을 거친 후, 골

다공증을 진단하는 문제에서 모델 성능을 최적화하기 위해 미세조정 학습을 진행하였다. 이는 모델이 어깨 X-ray 이미지에서 확인되는 골다공증의 특징을 더욱 정밀하게 인식하고 정확한 진단을 내릴 수 있도록 한다. RandomAug의 강도를 6, weight decay 0.05, layer-wise learning rate decay 0.55로 설정하였다. 사전 학습된 모델의 가중치를 최대한 보존하면서 미세 조정할 수 있도록 학습률을 $6.25e-05$ 로 설정하였다. 이는 기존에 학습된 네트워크 구조와 파라미터에 큰 변화를 주지 않으면서도, 새로운 데이터에 효과적으로 적응할 수 있도록 하기 위함이다.

4. 실험 결과

Model	Learning Method	ACC	AUC
ViT-small	흉부 x-ray MAE	0.756	0.842
Densenet121	흉부 x-ray MAE	0.727	0.819
Densenet121	흉부 x-ray MAE 어깨 x-ray 회전 proxy	0.744	0.819
Densenet121	흉부 x-ray MAE 어깨x-ray MAE(random crop)	<u>0.768</u>	<u>0.848</u>
Densenet121	흉부 x-ray MAE 어깨 x-ray MAE(80% center crop)	<u>0.768</u>	<u>0.848</u>

표1. 사전학습 시나리오와 진단 모델 종류(Densenet121, ViT-small)에 따른 어깨 x-ray 골다공증 진단 모델의 ACC, AUC

흉부 X-ray 이미지에서 MAE를 진행한 후 바로 골다공증 진단을 진행한 경우, ViT-small이 Densenet121에 비해 높은 성능을 보였다.

두 단계의 자기지도학습을 진행한 모델이 한 번 진행한 모델보다 전체적으로 높은 성능을 보였다. 회전 proxy task 방법을 사용한 경우, 흉부 X-ray MAE만을 진행했을 때와 비교했을 때 AUC 상의 변화는 없지만 더 높은 accuracy를 보였다. 어깨 X-ray MAE를 추가적으로 진행한 경우, 회전 proxy task를 진행했을 때보다 accuracy 0.768, ROC-AUC 0.848로 더 높은 성능을 보였다. Random crop과 Center crop간에는 성능 간에 큰 차이를 보이지 않았다.

5. 결론 및 향후연구

흉부 X-ray 이미지에서 MAE를 통한 사전학습 후 어깨 X-ray 이미지를 활용한 추가 사전학습을 진행했을 때, 모델의 골다공증 진단 성능이 향상된 것을 확인할 수 있다. 이는 흉부 X-ray 이미지에서는 학습할 수 없는 어깨 X-ray 이미지에서의 특성들을 추가적으로 모델이 학습할 수 있었기 때문인 걸로 판단된다.

두 단계의 사전학습을 거친 모델이 단일 사전학습 모델보다 높은 성능을 보였으며, 회전 proxy task를 진행한 경우보다 MAE를 진행한 경우 더 높은 성능을 보였다. 이는 어깨 X-ray

이미지의 회전을 통해 추출할 수 있는 특성이 골다공증을 세밀하게 진단하는 데에는 크게 활용되지 않았음을 의미한다.

이 연구는 흉부 X-ray 이미지에서의 MAE와 어깨 X-ray 이미지에서의 추가적인 자기지도 학습 방법은 의료 영상 분석에서 매우 유망한 접근법임을 보여준다. 특히 골다공증 진단과 같은 특정 질환 진단에서 더 나은 모델 성능을 가지도록 할 수 있다. 의료 분야를 넘어, 학습 데이터 수가 부족한 경우에서의 모델 학습과 causal 추론 능력이 중요한 분야에서의 모델 학습에서 본 논문은 중요한 기여를 할 수 있다.

향후 연구로는 X-ray 이미지를 넘어 다양한 의료 이미지 분야로 확장할 계획이다. 이를 통해 다양한 유형의 의료 이미지에 적용 가능한, 보다 범용적인 진단 모델을 개발하여 진단 정확성을 더욱 향상시키는 것을 목표로 한다. 추가적으로, 모델의 Causal 추론 능력에 대한 정량적 평가를 통해 더욱 심도 있는 연구를 할 수 있을 것으로 기대된다.

참 고 문 헌

- [1] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In 2017 IEEE International Conference on Computer Vision (ICCV) (pp. 618-626). IEEE.
- [2] He, K., Chen, X., Xie, S., Li, Y., Dollár, P., & Girshick, R. (2022). Masked Autoencoders Are Scalable Vision Learners. In 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) CVPR Computer Vision and Pattern Recognition (CVPR) (pp. 15979-15988). IEEE/CVF.
- [3] Xiao, J., Bai, Y., Yuille, A., & Zhou, Z. (2023). Delving into Masked Autoencoders for Multi-Label Thorax Disease Classification. In 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) (pp. 3577-3589). IEEE/CVF.