



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

의학 석사 학위 논문

인공 지능을 이용한 방사선 촬영 영상
분석을 통한 비골 골절의 진단과
임상적 적용

Diagnosis of nasal bone fracture and clinical
application with nasal bone x ray using artificial
intelligence

울 산 대 학 교 대 학 원

의 학 과

민 재 청

인공 지능을 이용한 방사선 촬영 영상
분석을 통한 비골 골절의 진단과
임상적 적용

지도교수 최종우

이 논문을 의학석사 학위 논문으로 제출함

2024년 2월

울산대학교 대학원
의학과
민재청

민재청의 의학석사 학위 논문을 인준함

심사위원 최 종 우 (인)

심사위원 오 태 석 (인)

심사위원 박 창 식 (인)

울 산 대 학 교 대 학 원

2024 년 2 월

목 차

국문요약.....	1
표차례.....	4
그림차례.....	5
서론.....	6
방법 및 모델.....	8
결과.....	12
고찰.....	14
결론.....	17
참고문헌.....	18
표.....	19
그림.....	20
영문요약.....	25

국 문 요 약

연구 목적

비골 골절의 진단을 위해서는 이 중 영상 검사가 중요한 역할을 하는데 이 중 X 선 검사는 검사 시행이 용이하며 방사선 노출 위험이 적고 80% 이상의 비교적 높은 민감도를 나타내는 것으로 알려져 있다. 그렇지만 숙련되지 않은 의사들에게 있어 X 선 검사를 통한 골절 진단은 어려운 과정일 수 있다. 근래에 의료 영상 진단 분야에서 기존의 방법론에 더해 최첨단 기술을 융합하며 새로운 변화를 나타내고 있다. 이러한 변화의 흐름에서 인공 지능 (Artificial intelligence) 은 정확하고 효율적으로 진단하는데 있어 중요한 역할을 하고 있다. 본 연구에서는 사전 학습 (pre training) 기법을 이용하여 X 선 검사를 통해 보다 정확하고 민감하게 비골 골절을 진단할 수 있는 방법을 마련하고자 한다. 이 기법을 최초로 X 선을 이용한 비골 골절 진단에 적용함으로써 높은 정확도의 진단 기법을 확립하여 숙련되지 않은 의사의 환자 진료에 도움을 줄 수 있을 것으로 생각된다.

연구 방법

Dataset

본 연구에서 사용된 데이터셋은 2010년 1월부터 2023년 3월 까지 비골 골절 의심하에 방사선 영상을 촬영한 총 4,058 명의 환자로부터 8,116 장의 측면 두부계측방사선 촬영 영상(Lateral cephalogram)으로 구성되었다. 데이터셋은 랜덤 샘플링(Random sampling) 기법을 통해 7:1:2의 비율로 훈련(Training), 검증(Validation), 평가(Test) 데이터셋으로 나뉘었다.

Imaging Crop

자체 개발한 기준점 설정 프로그램을 이용하여 기준점을 설정하고 기준점 사이에 중간 지점을 표시하여, 이를 기준으로 이미지는 잘라준다.

학습 구조

(a) 사전 학습 (Pre-training)

과제 : 골절선 분할 (Fracture line segmentation)

이를 위해, 코뼈 골절선(Nasal bone fracture line)의 관심 영역(Region of interest, ROI)은 숙련된 임상의로 인해 라벨링(Labeling) 되었다. 또한, 사전 학습에 사용된 영상 360 장은 코뼈 골절 분류의 왜곡된 평가를 방지하기 위해 전체 데이터셋 중 훈련과 검증 데이터셋으로만 구성되었다.

(b) 훈련 (Training)

과제 : 골절 분류 (Fracture classification)

Nam et al(2022)에서 제안된 모델 구조에 ImageNet Pretrained weight 을 사용하여 Baseline 을 설정였으며, 본 연구에서는 Figure 4-1 의 사전학습에서 학습된 Efficientnet backbone 의 인코더(Encoder)를 Figure 4-2 의 골절 분류 모델에 Transfer(모델의 가중치 적용) 하여 진행하였다.

(c) 추론 (Inference)

과제 : 골절 분류

Nam et al(2022)에 기반하여 영상 View 를 바꿔가며 2 개의 분류 모델에 각각 입력하였다.

연구 결과

본 연구에서 모델의 정량적 평가를 위해, 민감도(Sensitivity), 특이도(Specificity), 정확도(Accuracy), AUC(Area Under the Receiver Operating Characteristic Curve) 수치를 이용하였다.

민감도와 특이도, 정확도는 Tuning set 에서 참양성률(True positive rate)과 위양성률(False positive rate)의 차이가 가장 큰 임계값을 구하고 이를 Test set 에 대한 최적 임계값(Optimal threshold)로 적용함으로써 계산되었다. Baseline(Nam et al, 2022)에서 0.288 과 Our method 에서 0.667 의 최적 임계값을 나타내었다. 256x256 이미지 크기일 때, Baseline 모델에서는 70.1 %의 정확도, 63.2%의 민감도, 77.5%의 특이도를 나타내었으며, Our method 에서는 72.4%의 정확도, 65.2%의 민감도, 80.0%의 특이도를 나타내었다. 448x448 이미지 크기일 때, Baseline 모델에서는 72.9 %의 정확도, 71.1%의 민감도, 74.7%의 특이도를 나타내었으며, Our method 에서는 73.0%의 정확도, 64.0%의 민감도, 82.5%의 특이도를 나타내었다. Baseline 과 Our method 결과의 비교 시, Our method 에서 Baseline 보다 더 높은 특이도를 보이지만, 민감도와

AUC 는 상대적으로 낮은 수치를 보인다. 또한, 영상에서 Fracture 예측에 대한 관심 영역을 가시화 하기 위해 Grad-CAM 을 이용하였다. Grad-CAM 분석 결과, 참양성(True Positive)와 위양성(False Positive) 영상에서 모두 Nasal bone 영역에 대한 관심도가 높은 것으로 나타난다.

결론

본 연구를 통해 경험이 적은 의사가 방사선 검사 만으로도 비골 골절 환자의 진단을 용이하게 할 수 있도록 도와줄 수 있는 방법을 도출하였다. 임상적 소견과 인공 지능을 통해 얻은 골절 의심 소견 이미지를 통해 더 정확한 진단을 이룩할 수 있다.

표 차례

Table 1. 훈련, 검증, 평가 데이터셋 별 나이, 성별, 골절 비율 비교

Table 2. 256x256 Image 크기에서 기존 모델과 연구 모델 간 정확도, 민감도, 특이도, AUC 곡선 아래 면적 비교

Table 3. 448x448 Image 크기에서 기존 모델과 연구 모델 간 정확도, 민감도, 특이도, AUC 곡선 아래 면적 비교

그림차례

Figure 1. 본 연구에서 제외된 영상 사례

Figure 2. Imaging crop 과정

Figure 3-1. 골절선 분할을 위한 사전 학습을 적용한 연구 모델

Figure 3-2. 골절 분류를 위한 학습을 적용한 연구 모델

Figure 3-3. 골절 분류를 위한 추론 연구 모델

Figure 4. 기존 모델과 연구 모델을 256x256 Image 에 적용한 혼동행렬(Confusion Matrix)

Figure 5. 기존 모델과 연구 모델을 256x256 Image 에 적용한 AUC 곡선

Figure 6. 기존 모델과 연구 모델을 448x448 Image 에 적용한 혼동행렬(Confusion Matrix)

Figure 7. 기존 모델과 연구 모델을 448x448 Image 에 적용한 AUC 곡선

Figure 8. 골절이 있는 이미지를 골절이 있다고 적절하게 결론을 내린 사례에 대한 Grad-CAM

Figure 9. 골절이 없는 이미지를 골절이 있다고 잘못 결론을 내린 사례에 대한 Grad-CAM

Figure 10. 능동적 과정 학습을 적용한 가상 모델

I. 서론

비골은 안면 부위에서 가장 돌출된 부분으로 골절에 특히나 취약한 부위이다. 안면 골절의 절반 이상을 차지하고, 전체 골절의 1/3 이상을 차지하고 있다.[1, 2] 비골 골절을 진단하기 위해서는 환자 문진과 신체 진찰, 그리고 x 선 검사 및 CT 등의 영상 검사가 필요하다.

비골 골절의 진단을 위해서는 이 중 영상 검사가 중요한 역할을 하는데 특히 CT가 검사가 진단의 정확도가 가장 높은 것으로 알려져 있다.[3] 그렇지만 CT 검사를 시행하기 어려운 경우가 있고, 방사선 노출의 위험 또한 배제할 수 없다. 반면, X 선 검사는 검사 시행이 용이하며 방사선 노출 위험이 적고 80% 이상의 비교적 높은 민감도를 나타내는 것으로 알려져 있다. 그렇지만 숙련되지 않은 의사들에게 있어 X 선 검사를 통한 골절 진단은 어려운 과정일 수 있다.

근래에 의료 영상 진단 분야에서 기존의 방법론에 더해 최첨단 기술을 융합하며 새로운 변화를 나타내고 있다. 이러한 변화의 흐름에서 인공 지능 (Artificial intelligence) 은 정확하고 효율적으로 진단하는 데에 있어 중요한 역할을 하고 있다. 최근에는 합성곱 신경망 (convolutional neural network) 기법이 CT, MR, X 선 검사 등 다양한 영상 검사의 분석에 사용되고 있다. CT 검사를 이용한 비골 골절 판독을 인공지능을 이용하여 보조해 정확도를 높인 연구가 보고 되었고 [4] 흉부, 대퇴골, 두개골 골절 등을 X 선 검사 및 인공 지능을 이용해 진단한 결과도 보고 되었다.[5, 6] 비골 골절을 X 선 검사와 인공지능을 이용한 분석 또한 이루어져 85%의 정확도를 보이는 것으로 보고 되었다. [7] 제한된 환경에서도 간단하게 비골 골절을 진단하기 위해서는 X 선 검사만을 이용한 비골 골절 진단이 용이한데 이를 위해서는 이전보다 높은 정확도가 필요할 것으로 사료된다.

본 연구에서는 strong labelling 을 통한 사전 학습(pre training) 기법을 이용하여 X 선 검사를 통해 보다 정확하고 민감하게 비골 골절을 진단할 수 있는 방법을 마련하고자 한다. 이 기법을 최초로 X 선을 이용한 비골 골절 진단에 적용함으로써 높은 정확도의

진단 기법을 확립하여 숙련되지 않은 의사의 환자 진료에 도움을 줄 수 있을 것으로
생각된다.

II. 방법 및 모델

Dataset

본 연구에서 사용된 데이터셋은 2010년 1월부터 2023년 3월 까지 비골 골절 의심하에 방사선 영상을 촬영한 총 4,058 명의 환자로부터 8,116 장의 측면 두부측방방사선 촬영 영상(Lateral cephalogram)으로 구성되었다. 데이터셋에 포함된 영상은 양측에서 촬영된 영상이 모두 존재하는 환자의 영상에 한하며, 만 10세 이하의 환자에게 촬영된 영상, 해상도(Resolution)가 낮아 해부학적 구조(Anatomical structure)를 확인하기 어려운 영상과 앞선 예외 처리된 데이터셋 중 단일 영상(Single view)는 실험에서 제외 되었다. 환자의 코뼈 골절(Nasal bone fracture) 존재 유무에 대한 판단은 후향적 전자의무기록 분석을 통해 이루어졌다. 데이터셋은 랜덤 샘플링(Random sampling) 기법을 통해 7:1:2의 비율로 훈련(Training), 검증(Validation), 평가(Test) 데이터셋으로 나뉘었다. 훈련, 검증, 평가 데이터셋은 각각 2,839, 405, 814 명의 환자로부터 수집된 영상으로 구성되었으며, 해당 데이터셋의 특성은 표 1에 제시되었다.

Imaging Crop

원본 이미지에 대해 기준점 (Landmark) 설정을 위해 Gamma correction (사람의 눈은 어두운 부분의 변화는 크게 느끼고 밝은 부분의 변화는 작게 느껴 이를 보정)을 시행한다. 자체 개발한 기준점 설정 프로그램을 이용하여 기준점을 설정하고 기준점 사이에 중간 지점을 표시하여, 이를 기준으로 이미지는 잘라준다. [8] (기준점 : 연부 조직 비근점, 비주)

학습 구조

Efficientnet는 MBConv와 SENet으로 구성된 베이스 모델 구조를 가지며, 모델의 너비(Width), 깊이(Depth), 해상도(Resolution) 계수를 균형 있게 스케일(Scaling) 하는 Compound scaling method를 통해 높은 성능을 도출 하는 것으로 알려져 있다. 본

연구에서는 모델의 파라미터(Parameter) 수가 가장 큰 Efficientnet b7 을 Backbone network 로 선정하였다.

사전학습의 경우, Efficientnet b7 을 인코더(Encoder)의 Backbone 으로 한 U-Net 모델을 학습하여 골절선 분할 (Fracture line segmentation)을 진행하였다.

(d) 사전 학습 (Pre-training)

과제 : 골절선 분할 (Fracture line segmentation)

사전 학습(Pre-training)을 위한 데이터셋은 243 명의 환자로부터 수집된 360 장의 영상으로 구성되었다. 사전 학습은 코뼈 골절선(Nasal bone fracture line)을 분할(Segmentation)하고, 이를 코뼈 골절 분류에 있어 초기 가중치로 사용함으로써 효율적이고 효과적인 학습을 하는 것을 목표로 한다. 이를 위해, 코뼈 골절선(Nasal bone fracture line)의 관심 영역(Region of interest, ROI)은 숙련된 임상의를 의해 라벨링(Labeling) 되었다. 또한, 사전 학습에 사용된 영상 360 장은 코뼈 골절 분류의 왜곡된 평가를 방지하기 위해 전체 데이터셋 중 훈련과 검증 데이터셋으로만 구성되었다. 최종적으로 이를 8:2 의 비율로 나누어 사전학습을 위한 훈련과 검증 데이터셋을 구성하였다.

(e) 사전 학습 : 구현 상세 기술

U-Net Backbone: Efficientnet b7

골절선 관심 영역을 기준으로 높이, 너비가 10cm 가 되도록 부분 추출(Patch extraction) 골절선 관심 영역에 5x5 중심(kerne)의 확장(dilation) 적용

이미지 크기 : 448x448 Batch size: 4 Epoch: 100 Loss function: Dice loss

데이터 확대(Data augmentation): Translation, Scaling, Rotation, Gaussian Noise

Optimizer: Adam, learning rate $1e-4$, weight decay $5e-4$

Clahe 적용 (contrast limited adaptive histogram equalization): 영상의 대조도를 조정

Tuning set 에서 Dice Loss 가 가장 낮을 때 모델 사용

(f) 훈련 (Training)

과제 : 골절 분류 (Fracture classification)

Figure 3-2 에 도시된 바와 같이, 골절 분류를 위한 모델은 *Nam et al(2022)*에 기반하여 2 개의 Efficientnet b7 Network 와 MLP(Multi-Layer Perceptron)으로 구성되었다. 실험에 사용된 Multi view 의 영상은 각각 서로 다른 Efficientnet b7 Network 로 입력되어진다.

다음으로 모델은 각 Efficientnet b7 Network 에서 얻어진 2,560 특징(Features)을 5,620 특징으로 접합(Concatenation) 한 후, 이 features 가 다층 퍼셉트론(Multi Layer Perceptron)을 거침으로서 정상과 골절에 대한 출력 확률값(Output probability)을 출력한다.

위와 같이, *Nam et al(2022)*에서 제안된 모델 구조에 ImageNet Pretrained weight 을 사용하여 Baseline 을 설정였으며, 본 연구에서는 Figure 3-1 의 사전학습에서 학습된 Efficientnet backbone 의 인코더(Encoder)를 Figure 3-2 의 골절 분류 모델에 Transfer(모델의 가중치 적용) 하여 진행하였다.

그 후, Total dataset 의 Train set 을 통해 골절 분류 모델을 Full fine-tuning 하였으며, Tuning set 을 통해 분류 기준(Cut-off)을 위한 최적 역치(Optimal threshold) 선정, 모델의 학습 종료 및 선택(Selection) 등을 진행 하였다.

(g) 훈련 : 구현 상세 기술

Model Backbone: Efficientnet b7

이미지 크기: 448x448, Batch size: 8, Epoch: 100, Loss function: Cross entropy loss

데이터 확대(Data augmentation): Translation, Scaling, Rotation , Gaussian Noise

Optimizer: Adam, learning rate $1e-4$, weight decay $5e-4$

Clah 적용 (contrast limited adaptive histogram equalization): 영상의 대조도를 조정

Tuning set 에서 Cross Entropy Loss 가 가장 낮을 때 모델 사용.

역치(Threshold): Tuning set 의 Optimal threshold

(h) : 추론 (Inference)

과제 : 골절 분류

Figure 3-2 에서 학습된 분류 모델을 전체 데이터셋 중 시험셋(Test set)으로 추론하였다.

추론 시, *Nam et al(2022)*에 기반하여 영상 View 를 바꿔가며 2 개의 분류 모델에 각각 입력하였다.

정상/골절 분류를 위한 최종 출력 확률값(Final output probability)은 각 분류 모델에서 출력된 출력 확률값들의 평균 값으로 하였다.

Final output probability 값을 Tuning set 에서 계산된 이상적 기준 역치(Optimal Cut-off Threshold) 값을 기준으로 정상과 골절을 판단하였다.

III. 결과

본 연구에서 모델의 정량적 평가를 위해, 민감도(Sensitivity), 특이도(Specificity), 정확도(Accuracy), AUC(Area Under the Receiver Operating Characteristic Curve) 수치를 이용하였다.

민감도는 모델이 실제 Fracture 가 있는 환자를 Fracture 가 있다고 판단할 확률, 특이도는 실제 정상(Normal)인 환자를 정상이라고 판단할 확률, 정확도는 Fracture 와 정상 환자를 모두 정확히 판단할 확률을 의미한다. AUC 는 다양한 임계값(Threshold)에서 모델의 분류 성능을 도시화하는 ROC (Receiver Operating Characteristic) curve 의 면적을 의미한다. 이 중 민감도와 특이도, 정확도는 Tuning set 에서 참양성률(True positive rate)과 위양성률(False positive rate)의 차이가 가장 큰 임계값을 구하고 이를 Test set 에 대한 최적 임계값(Optimal threshold)로 적용함으로써 계산되었다.

Baseline(Nam et al, 2022)에서 0.288 과 Our method 에서 0.667 의 최적 임계값을 나타내었다. 이를 통해, Table2 에 도시된 바와 같이, 256x256 이미지 크기일 때, Baseline 모델에서는 70.1 %의 정확도, 63.2%의 민감도, 77.5%의 특이도를 나타내었으며, Our method 에서는 72.4%의 정확도, 65.2%의 민감도, 80.0%의 특이도를 나타내었다. Table3 에 도시된 바와 같이 448x448 이미지 크기일 때 Baseline 모델에서는 72.9 %의 정확도, 71.1%의 민감도, 74.7%의 특이도를 나타내었으며, Our method 에서는 73.0%의 정확도, 64.0%의 민감도, 82.5%의 특이도를 나타내었다.

256x256 이미지 크기 및 448x448 이미지 크기에서 골절에 대한 예측 결과와 실제 결과를 비교하기 위한 혼동행렬 (Confusion matrix) 및 AUC 곡선은 Figure 4 와 6 에 도시되었다. 다음으로 다양한 임계값에서 모델의 분류 성능을 평가하는 ROC curve 는 이미지 크기에 따라 Figure 5 와 7 에 도시되었다. ROC curve 의 AUC 는 256x256 이미지 크기에서는 Baseline 에서 80.1%, Our method 에서 79.1% , 448x448 이미지 크기에서는 Baseline 에서 76.0%, Our method 에서 79.0% 수치를 나타내었다.

Baseline 과 Our method 결과의 비교 시, Our method 에서 Baseline 보다 더 높은 특이도를 보이지만, 민감도와 AUC 는 상대적으로 낮은 수치를 보인다.

또한, 영상에서 Fracture 예측에 대한 관심 영역을 가시화 하기 위해 Grad-CAM 을 이용하였다. 적색 영역은 Fracture 판단에 있어 관심이 높은 영역, 청색 영역은 관심이 낮은 영역을 의미한다. Grad-CAM 분석 결과, 참양성(True Positive)와 위양성(False Positive) 영상에서 모두 Nasal bone 영역에 대한 관심이 높은 것으로 나타난다. Grad-CAM 은 참양성과 위양성 군 각각을 Figure 8 과 9 에 도시하였다.

IV. 고찰

영상 검사를 통해 숙련된 의사의 진료로 적절한 진단을 내릴 수 있다면 이상적이겠으나 제한된 환경에서 진료 시에 전문적인 도움을 받을 수 있다면 환자 진료 및 치료에 기여할 수 있을 것이다. 이러한 측면에서 의료 진단 분야에서 인공지능의 역할이 대두되고 있으며, 다양한 연구에서 골절 진단에서 인공지능의 기여도를 보여주었다. Sato 등은 인공지능을 통해 X선 검사로 효과적인 고관절 골절 진단능을 검증해내었으며[5], Kuo 등은 여러 부위의 골절에서 인공지능이 진단능을 높일 수 있다는 사실을 증명하였다.[6] 비골은 사지와는 달리 해부학적 봉합선과 구조적 특성 때문에 인공지능으로 재현하기 어려운 부분이 있다. 그렇지만 비골 골절에 대해서도 인공지능을 이용하여 진단능을 높이고자 연구한 사례들이 있었다. Seol 등은 삼차원 합성곱 신경망을 이용하여 X 선 검사를 통해 비골 골절을 진단하고자 하였다. 그 결과 83.1%의 민감도, 83.7%의 특이도 및 83.3%의 정확도를 나타내었다.[9] Yang 등은 CT 검사에 딥 러닝을 모델을 결합하여, CT 단독 검사보다 높은 민감도, 특이도 및 정확도의 결과를 보여주었다.[4, 8, 10]

본 연구에서는 의사의 숙련도와 관계 없이 진단능을 높일 수 있도록 새로운 기법의 인공지능 훈련을 기반으로 단순 X 선 검사만을 통하여 높은 정확도로 비골 골절을 진단할 수 있는 시스템을 개발하고자 하였다.

진단의 정확도, 민감도, 그리고 특이도를 높이기 위하여 분석을 시행 하기 전 Imaging crop 기법의 효능을 검증하였다. Imaging crop 기법은 기준점(비근점, 비주)을 설정하여 분석하고자 하는 특정 부위만을 잘라내는 기법으로 Resnet 50, Densenet 121, Efficientnet v2 small 등 다양한 back bone 에서 Imaging crop 기법을 더하여 기존 방법 보다 향상된 정확도, 민감도, 특이도를 관찰할 수 있었다.

본 연구에서는 모델의 파라미터(Parameter) 수가 가장 큰 Efficientnet b7 을 Backbone network 로 선정하였다. 사전학습의 경우, Efficientnet b7 을 인코더(Encoder)의

Backbone 으로 한 U-Net 모델을 학습하여 Fracture line segmentation 을 진행하였다. 진단의 정확도, 민감도, 특이도 상승을 위해서 라벨링을 통한 사전 학습을 진행하였다. 숙련된 임상 의에 의해 진행된 사전 학습에서 Efficientnet backbone 의 Encoder 를 Fracture classification 모델에 Transfer 하여 진행하였다.

이미지 크기를 256x256 으로 진행하여 기존의 Efficientnet b7 을 본원에서 추출한 Data set 에 적용한 결과와 사전 학습을 적용한 본 연구의 결과를 비교한 결과, Table 2 에 도시된 것과 같이 더 높은 정확도, 민감도, 특이도를 보였으며 AUC curve 아래 면적 값 또한 0.76 에서 0.79 로 상승된 결과를 관찰할 수 있었다. 본 연구에서 총 814 개의 image 에서 589 개의 사례에서 정확한 진단(골절이 있는 것은 있다고 판단하고 골절이 없는 것은 없다고 판단한 사례)을 내려 72.4 의 정확도를 보였다.

Image size 를 키워 440x440 으로 진행하여 기존 Base line 방법과 본 연구의 사전 학습 방법을 더한 model 을 적용하여 비교한 결과 Table 3 에 도시된 것과 같이 두 model 모두에서 이전보다 높은 정확도, 민감도, 특이도, 그리고 높은 값의 AUC curve 아래 면적 값을 보였다. 이는 큰 Image 에서 골절선을 감지할 확률이 높아진 이유에 기인한 것으로 생각되며, 두 model 간의 유의한 값의 차이는 보이지 않았다. 본 연구에서 총 814 개의 image 에서 594 개의 사례에서 정확한 진단을 내려 73.0 의 정확도를 보였다.

Grad-CAM (Gradient - Class Activation Map)을 통해서 본 연구에 사용한 model 이 골절 부위를 인식한 것을 시각적으로 확인할 수 있다. Figure 8 에서는 골절이 있는 환자에서 골절 부위를 찾아 내어 골절이 있다고 결론을 내린 사례를 확인할 수 있다. Figure 9 에서는 골절이 없는 환자에서 골절이 있다고 오판을 한 사례를 확인할 수 있다. 이는 해부학적으로 본래 존재하는 봉합선이나, 예전에 골절이 생겼다가 치유된 골절선을 새로 생긴 골절로 잘못 결론을 내린 것으로 생각된다. 이러한 사례를 0%에 가깝게 만드는 것은 기술적으로 불가능하나 이러한 위양성 사례의 Grad-CAM 은 본 연구에서 사용한 모델이 비골의 어떤 영역에서 골절로 잘못 판단했는지 보여주며 임상적 연관성과 더해 실제 진단시에 정확도를 높이는 데에 도움을 줄 수 있을 것이다.

본 연구는 여러 제한점을 가지고 있다. 첫째로 추출된 Data set 자체가 X 선 검사만으로 골절을 판별하기 상대적으로 어려운 것으로 생각된다. 같은 Base line model 을 적용하였을 때 이전 연구에서 보다 본 연구에서 정확도, 민감도, 특이도, AUC 곡선 아래 면적 값이 유의하게 감소한 것으로 보였다. 이전에 사용했던 data set 에 사전 학습을 더한 본 연구의 model 을 적용하여 비교하면 다른 결과를 도출할 수 있을 것으로 생각된다. 두번째로 큰 이미지 크기 (448x448)에서 기존 모델과 유의한 차이를 보이지 못했다. 이러한 한계점을 개선하기 위해 Figure 10 에 도시된 것과 같이 능동적 과정학습 (Active curriculum learning)을 추후에 사전 학습에 적용하여 추후에 연구를 진행한다면 더 우수한 결과를 얻을 수 있을 것으로 생각된다.

V. 결론

본 연구를 통해 경험이 적은 의사가 방사선 검사 만으로도 비골 골절 환자의 진단을 용이하게 할 수 있도록 도와줄 수 있는 방법을 도출하였다. 이전 연구와 달리 Labelling 을 통한 사전 학습 모델을 적용하여 작은 사이즈의 이미지 크기에서 높은 정확도, 민감도, 특이도를 나타내었다. 임상적 소견과 인공 지능을 통해 얻은 골절 의심 소견 이미지를 통해 더 정확한 진단을 이룩할 수 있으며 추후에 능동적 과정 학습 모델을 적용한다면 더 개선된 결과를 나타낼 수 있을 것이다.

VI. 참고문헌

1. Hwang K, You SH, Kim SG, Lee SI. Analysis of nasal bone fractures; a six-year study of 503 patients. *J Craniofac Surg.* 2006 Mar;17(2):261-4. doi: 10.1097/00001665-200603000-00010. PMID: 16633172.
2. Li L, Zang H, Han D, Yang B, Desai SC, London NR. Nasal Bone Fractures: Analysis of 1193 Cases with an Emphasis on Coincident Adjacent Fractures. *Facial Plast Surg Aesthet Med.* 2020 Jul/Aug;22(4):249-254. doi: 10.1089/fpsam.2020.0026. Epub 2020 Apr 6. PMID: 32250646; PMCID: PMC7476382.
3. Hwang K, Jung JS, Kim H. Diagnostic Performance of Plain Film, Ultrasonography, and Computed Tomography in Nasal Bone Fractures: A Systematic Review. *Plast Surg (Oakv).* 2018 Nov;26(4):286-292. doi: 10.1177/2292550317749535. Epub 2018 Jan 9. PMID: 30450348; PMCID: PMC6236500.
4. Yang C, Yang L, Gao GD, Zong HQ, Gao D. Assessment of artificial intelligence-aided reading in the detection of nasal bone fractures. *Technol Health Care.* 2023;31(3):1017-1025. doi: 10.3233/THC-220501. PMID: 36442167.
5. Sato Y, Takegami Y, Asamoto T, Ono Y, Hidetoshi T, Goto R, Kitamura A, Honda S. Artificial intelligence improves the accuracy of residents in the diagnosis of hip fractures: a multicenter study. *BMC Musculoskelet Disord.* 2021 May 3;22(1):407. doi: 10.1186/s12891-021-04260-2. PMID: 33941145; PMCID: PMC8091525.
6. Kuo RYL, Harrison C, Curran TA, Jones B, Freethy A, Cussons D, Stewart M, Collins GS, Furniss D. Artificial Intelligence in Fracture Detection: A Systematic Review and Meta-Analysis. *Radiology.* 2022 Jul;304(1):50-62. doi: 10.1148/radiol.211785. Epub 2022 Mar 29. PMID: 35348381; PMCID: PMC9270679.
7. Nam, Y., Choi, Y., Kang, J. *et al.* Diagnosis of nasal bone fractures on plain radiographs via convolutional neural networks. *Sci Rep* **12**, 21510 (2022). <https://doi.org/10.1038/s41598-022-26161-7>.
8. Choi JW, Park H, Kim B S IH, Kim N, Kwon SM, Lee JY. Surgery-First Orthognathic Approach to Correct Facial Asymmetry: Artificial Intelligence-Based Cephalometric Analysis. *Plast Reconstr Surg.* 2022 Mar 1;149(3):496e-499e. doi: 10.1097/PRS.0000000000008818. PMID: 35196690.
9. Seol YJ, Kim YJ, Kim YS, Cheon YW, Kim KG. A Study on 3D Deep Learning-Based Automatic Diagnosis of Nasal Fractures. *Sensors (Basel).* 2022 Jan 10;22(2):506. doi: 10.3390/s22020506. PMID: 35062465; PMCID: PMC8780993.
10. Mohammadi A, Ghasemi-Rad M. Nasal bone fracture-ultrasonography or computed tomography? *Medical Ultrasonography.* (2011) ; 13: (4): 292-5. Epub 2011/12/02..

Table (표)

Table 1. 훈련, 검증, 평가 데이터셋 별 나이, 성별, 골절 비율 비교

Characteristics	Training set (n=2,839)	Tuning set (n=405)	Test set (n=814)
Age, mean±SD	37.5 (18.5)	39.5 (18.8)	36.9 (18.4)
Gender, n (%)			
Men	1,801 (63.4)	255 (63.0)	548 (67.3)
Women	1,038 (36.6)	150 (37.0)	266 (32.7)
Label, n (%)			
Normal	1,379 (48.6)	197 (48.6)	395 (48.5)
Fracture	1,460 (51.4)	208 (51.4)	419 (51.5)

Table 2. 256x256 Image 크기에서 기존 모델과 연구 모델 간 정확도, 민감도, 특이도, AUC 곡선 아래 면적 비교

Characteristics	Accuracy	Sensitivity	Specificity	AUC
Base line	70.1	63.2	77.5	76.0
Our method	72.4	65.2	80.0	79.0

Table 3. 448x448 Image 크기에서 기존 모델과 연구 모델 간 정확도, 민감도, 특이도, AUC 곡선 아래 면적 비교

	Accuracy	Sensitivity	Specificity	AUC
Base line (Nam et al. 2022)	72.9	71.1	74.7	80.1
Our method	73.0	64.0	82.5	79.1

그림 및 그래프

Figure 1. 본 연구에서 제외된 영상 사례



Figure 2. Imaging crop 과정

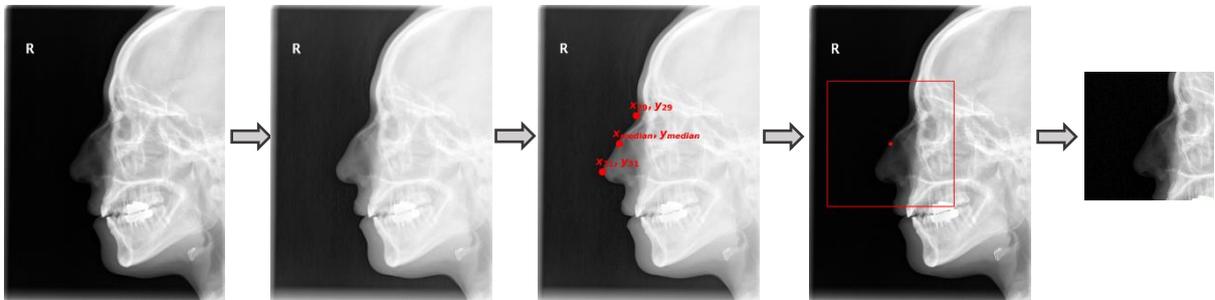
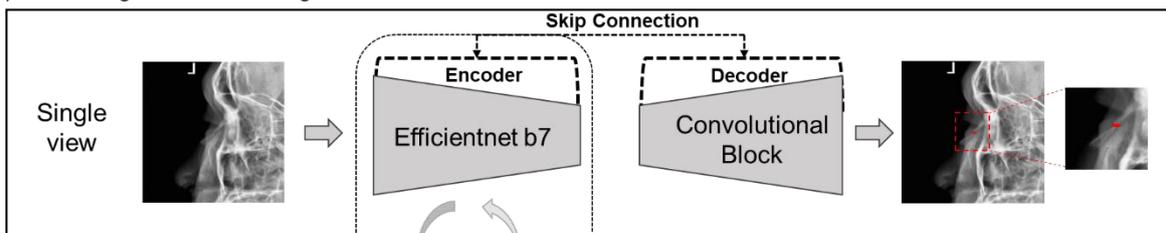


Figure 3-1. 골절선 분할을 위한 사전 학습을 적용한 연구 모델

Figure 3-2. 골절 분류를 위한 학습을 적용한 연구 모델

(a) Pre-training: Fracture line segmentation



(b) Training: Fracture classification

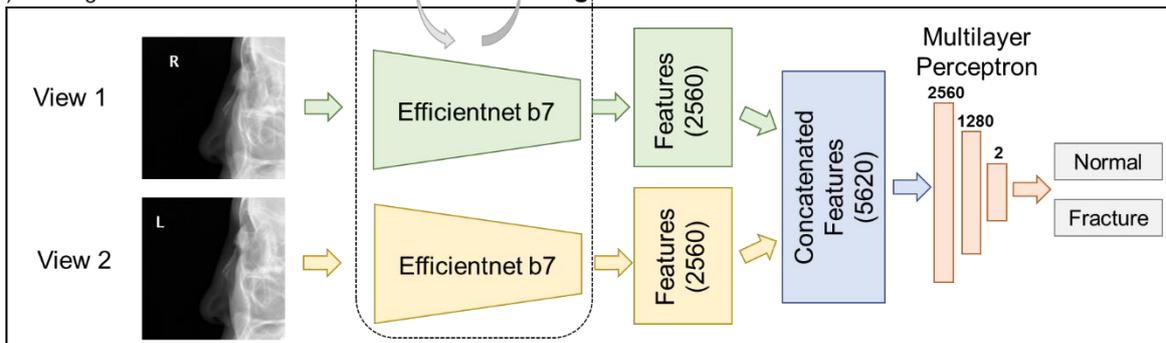


Figure 3-3. 골절 분류를 위한 추론 연구 모델

(c) Inference: Fracture classification

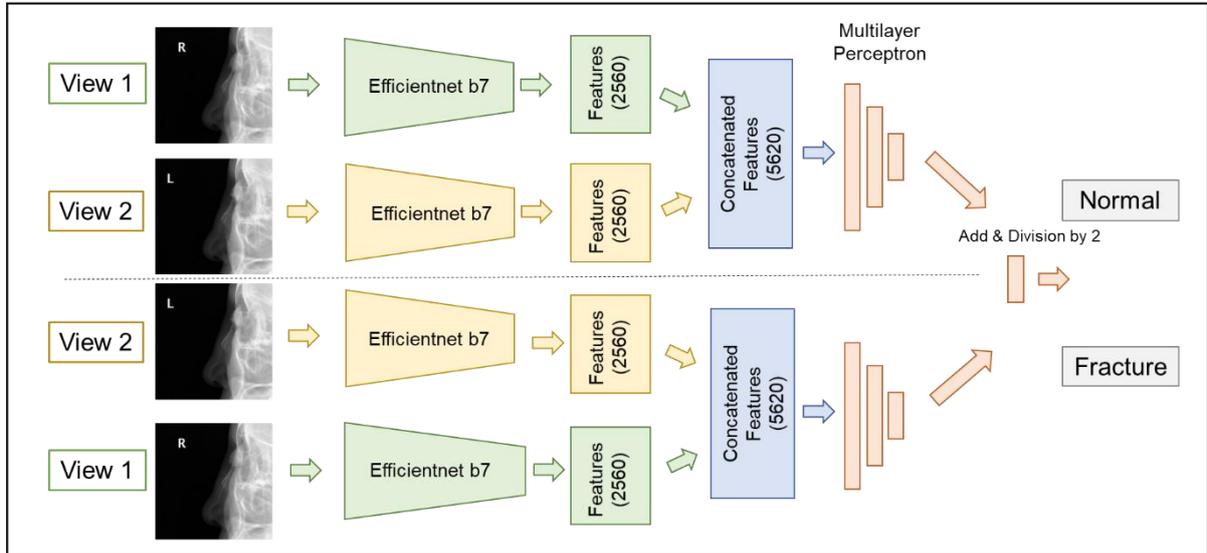


Figure 4. 기존 모델과 연구 모델을 256x256 이미지에 적용한 혼동행렬(Confusion Matrix)

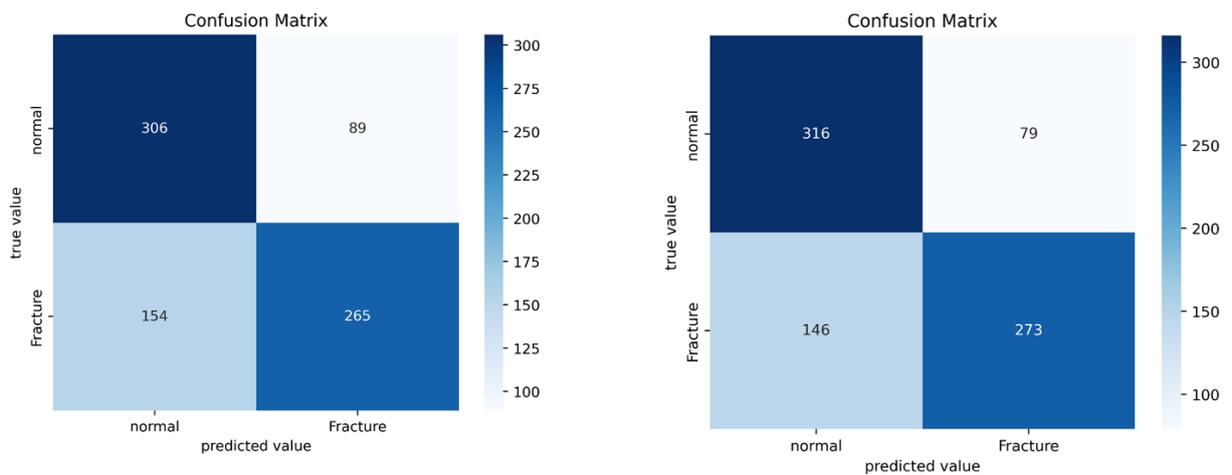


Figure 5. 기존 모델과 연구 모델을 256x256 이미지에 적용한 AUC 곡선

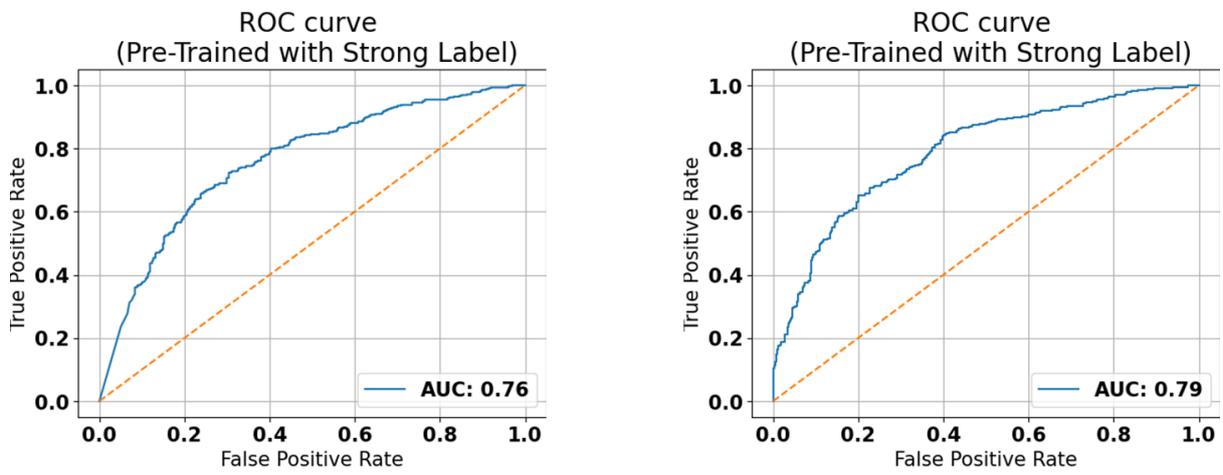


Figure 6. 기존 모델과 연구 모델을 448x448 이미지에 적용한 혼동행렬(Confusion Matrix)

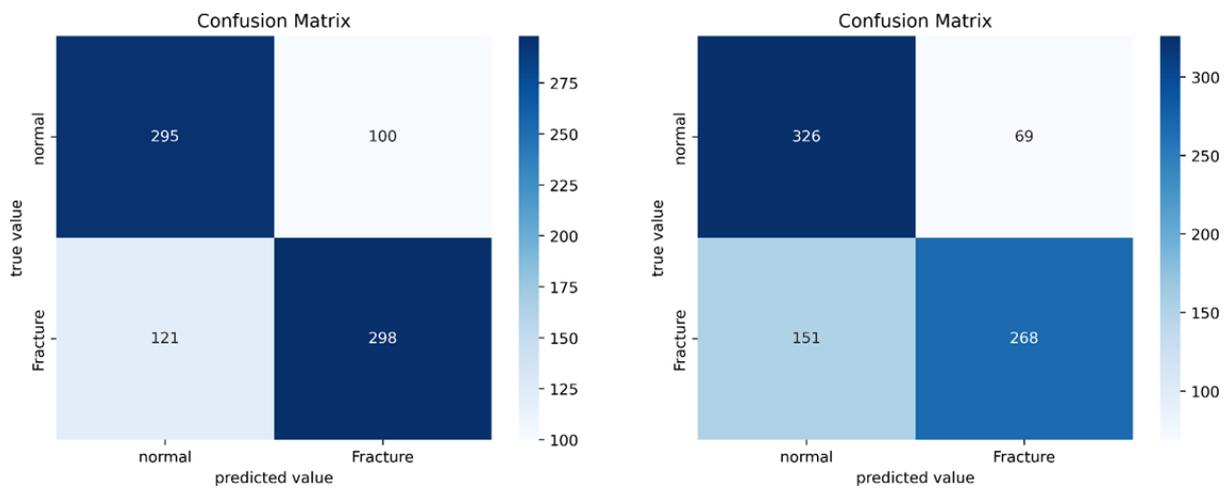


Figure 7. 기존 모델과 연구 모델을 448x448 이미지에 적용한 AUC 곡선

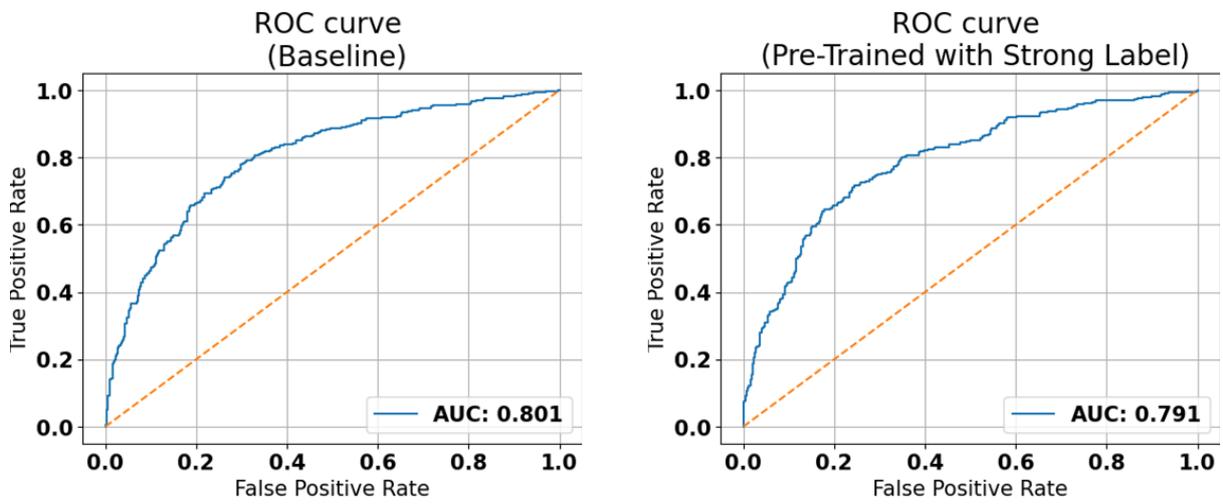


Figure 8. 골절이 있는 이미지를 골절이 있다고 적절하게 결론을 내린 사례에 대한 Grad-CAM

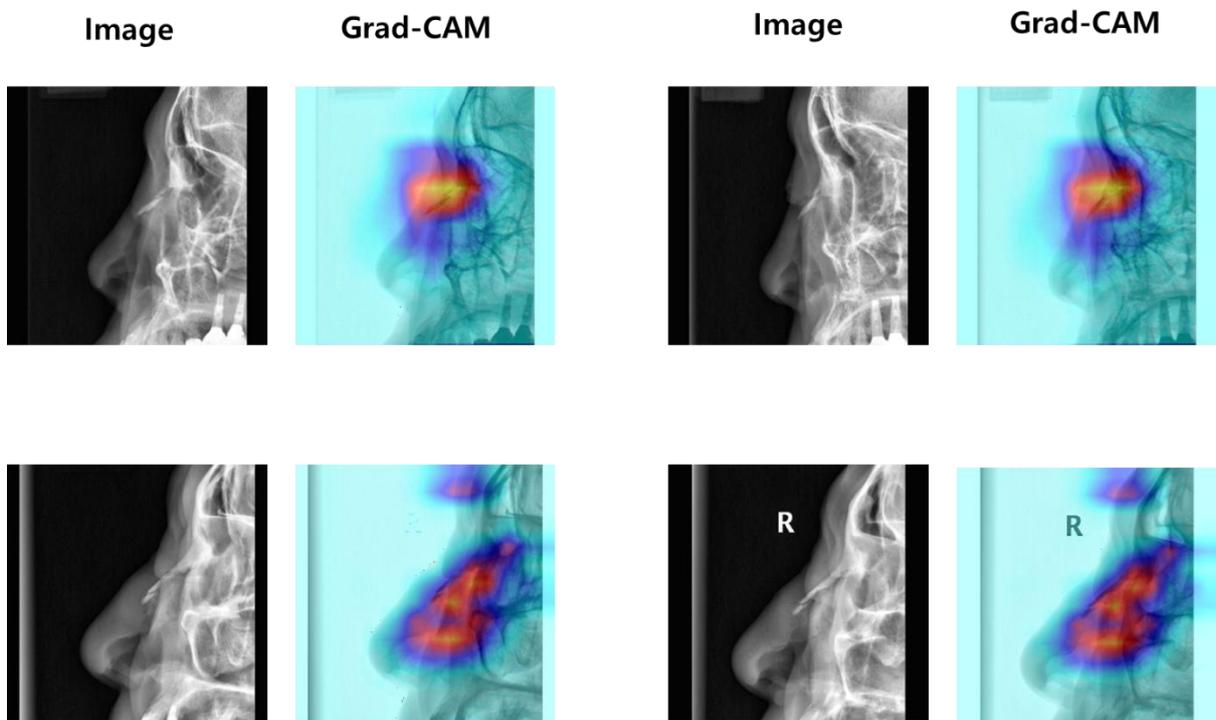


Figure 9. 골절이 없는 이미지를 골절이 있다고 잘못 결론을 내린 사례에 대한 Grad-CAM

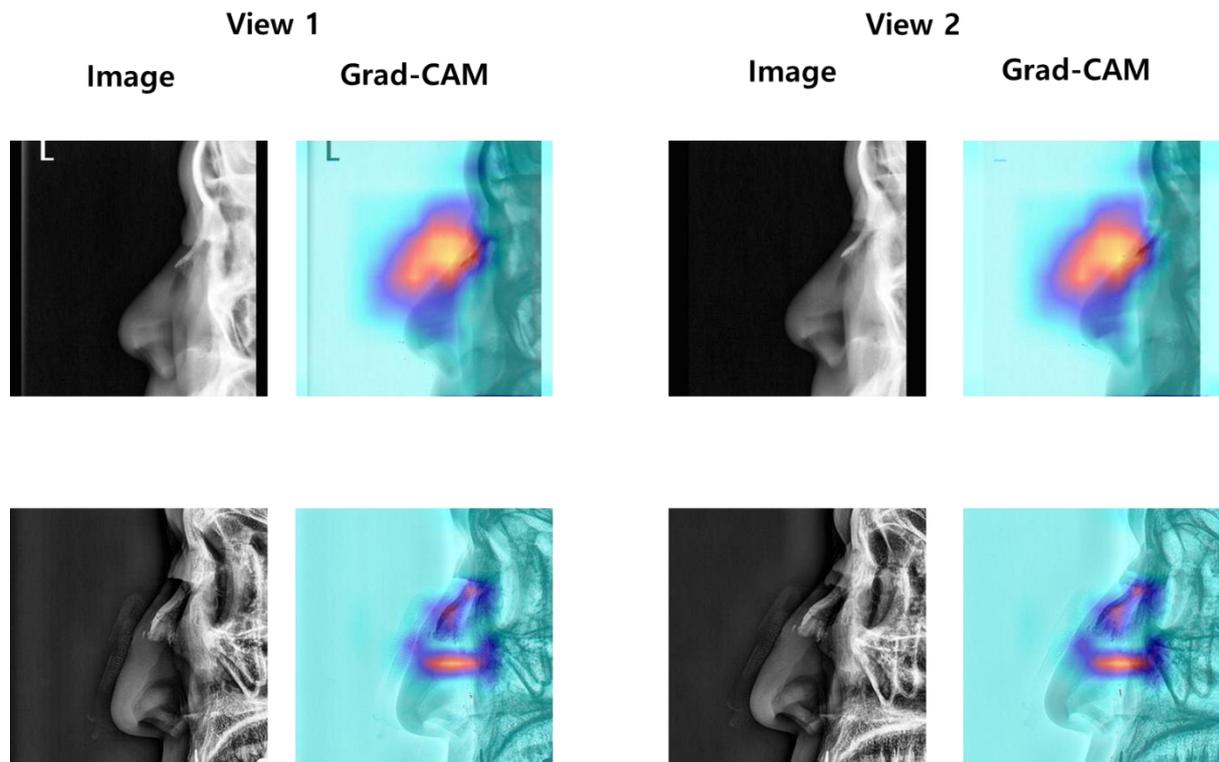
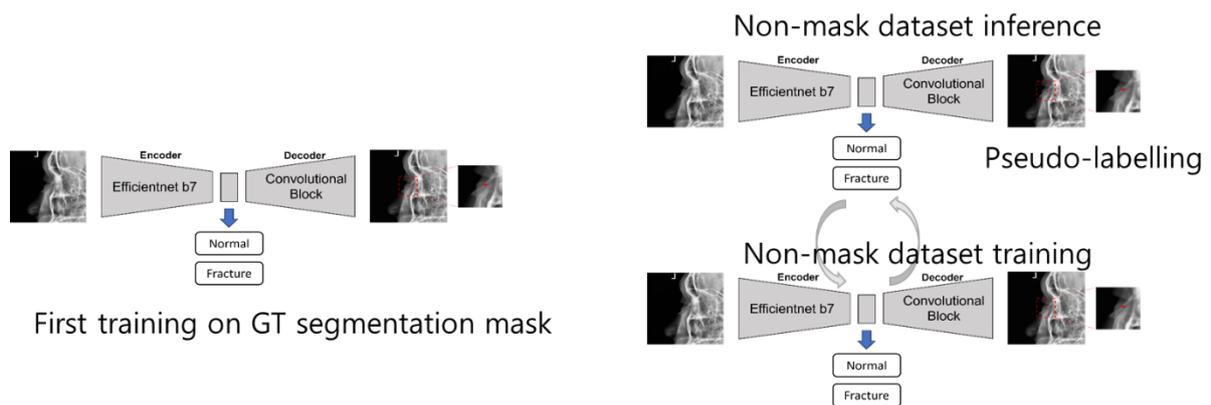


Figure 10. 능동적 과정 학습을 적용한 가상 모델



영문 요약

Introduction

For the diagnosis of nasal bone fractures, imaging modalities play an important role. Among these, X-ray imaging is known to be easy to perform, have a low risk of radiation exposure, and have a relatively high sensitivity of over 80%. However, it is difficult for unexperienced doctors to diagnose fractures with X-ray images only. Recently, the field of medical imaging diagnosis is showing new changes by converging cutting-edge technologies in addition to conventional methods. Within the current trajectory of transformation, artificial intelligence assumes a pivotal role in optimizing precision and efficacy in diagnostic processes. In this study, our objective is to devise a method for enhancing the accuracy and sensitivity of nasal bone fracture diagnoses via X-ray assessment, employing a pre-training technique.

Methods

Data set

The dataset utilized in this study comprises a total of 8,116 lateral cephalogram images, obtained through radiographic imaging for suspected fibula fractures from 4,058 patients, spanning from January 2010 to March 2023. The dataset was partitioned using a random sampling technique, allocating a ratio of 7:1:2 for training, validation, and testing, respectively.

Imaging crop

Utilizing an internally developed landmark-setting program, we establish reference points and demarcate intermediate points between these landmarks. Subsequently, the image is cropped based on these reference and intermediate points.

Training architecture

(a) Pre-training

Subject : Fracture line segmentation

For this purpose, the region of interest (ROI) pertaining to the nasal bone fracture line was labeled by a skilled clinical expert. Additionally, the set of 360 images utilized in pre-training was exclusively allocated to the training and validation datasets within the entire dataset, aiming to mitigate distorted assessments in nasal bone fracture classification.

(b) Training

Subject : Fracture classification

The model architecture proposed by Nam et al. (2022) was established as a baseline using ImageNet pretrained weights. In this study, the encoder of the Efficientnet backbone pretrained was transferred (i.e., applied as the model's weights) to the fracture classification model.

(c) Training

Subject : Fracture classification

Based on Nam et al. (2022), image views were systematically altered and input into two distinct classification models.

Results

In this study, for the quantitative evaluation of the model, sensitivity, specificity, accuracy, and the Area Under the Receiver Operating Characteristic Curve (AUC) were utilized. Sensitivity, specificity, and accuracy were calculated by determining the threshold with the largest difference between the true positive rate and false positive rate on the tuning set and applying it as the optimal threshold for the test set. The optimal thresholds were 0.288 in the baseline (Nam et al., 2022) and 0.667 in our method. For 256x256 image size, the baseline model showed an accuracy of 70.1%, sensitivity of 63.2%, and specificity of 77.5%, while our method demonstrated an accuracy of 72.4%, sensitivity of 65.2%, and specificity of 80.0%. For 448x448 image size, the baseline model exhibited an accuracy of 72.9%, sensitivity of 71.1%, and specificity of 74.7%, whereas our method showed an accuracy of 73.0%,

sensitivity of 64.0%, and specificity of 82.5%. When comparing the results between the baseline and our method, our method exhibited higher specificity, but sensitivity and AUC showed relatively lower values. Additionally, Grad-CAM was used to visualize regions of interest for fracture prediction in images. Grad-CAM analysis revealed high interest in the Nasal bone region in both true positive and false positive images.

Conclusion

This study derived a method to facilitate the diagnosis of patients with nasal bone fractures solely through radiographic examinations, enabling less experienced physicians to do so effortlessly. In contrast to prior research, the application of a pre-trained model through labeling demonstrated high accuracy, sensitivity, and specificity on smaller image sizes.