

# Pediatric Fractures

2025.11.03

김진규

---

# 목 차

## 1. 프로젝트 배경

문제제기

필요성

기대효과

## 3. 기술 사항

기술

## 4. 계획

## 2. 프로젝트 개요

진행계획

프로젝트 개요

주요기능

## 5. Q&A

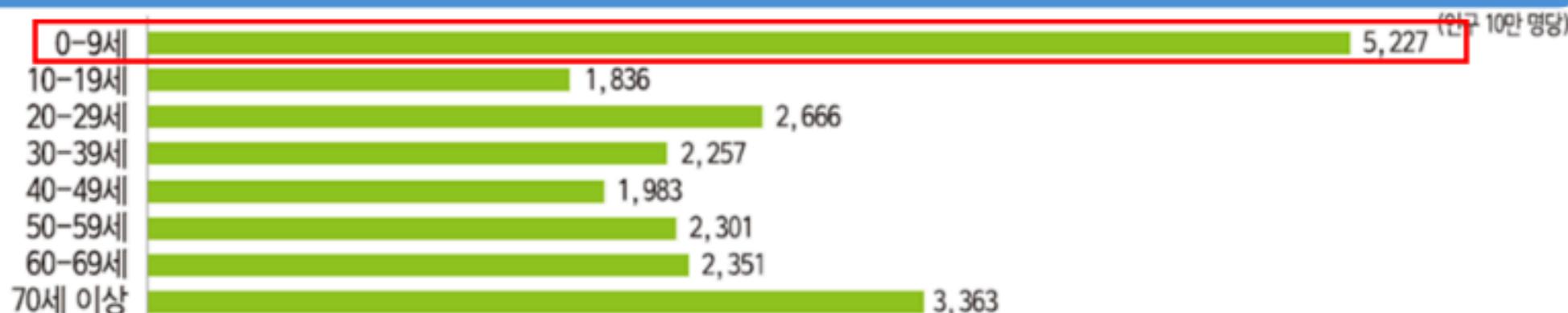
# **프로젝트 배경**

# 문제제기

## 연구 개발 과제의 중요성

소아 응급환자의 경우 X-ray 판독에서 골절 검출을 놓쳤을 시, 영구적 뼈 변형 또는 기능 손실을 초래할 수 있기에 신속하고 정확한 진단이 필요하다.

손상 응급실 방문 환자의 연령별 발생률(국가응급진료정보망)



소아는 전 연령 중 손상에  
의한 응급실 이용율이  
가장 많고 약 5-10%에서  
골절이 동반됨



신체 진찰만으로 골절을  
배제하기 힘든 경우가 많아  
x-ray 검사가 필수적임



뼈의 빠른 성장이 일어나는  
시기이므로 골절을 제때  
진단하지 않을 경우  
영구적인 뼈의 변형이나  
기능의 손실을 가져올 수 있음



성장 단계에 따른  
다양성으로 인해 판독이  
까다로워 최대 27%까지  
골절을 놓칠 수 있음

# 필요성

## 연구 개발 과제 개발 방향

의료진의 숙련도에 따라 여러 장의 촬영으로 방사선 노출이 증가 할 수 있어

**최소한의 방사선 촬영으로 신속하게 소아골절을 진단하는 모델 개발이 필요**

As-is

1. 숙련된 의료진은 최소 2개의 X-ray(AP/LAT)를 활용해 골절을 진단하며, 부위에 따라 3류 이상이 요구됨
2. 전공의·비숙련 의료진은 정확한 판단을 위해 추가 촬영하는 경우 많음
3. 환자의 방사선 노출 증가와 진단 지연 발생 가능
4. 소아 진료 취약 지역에서의 수요

To-be

1. 최소 영상(2장: AP/LAT)으로 골절 판독하는 AI 모델 개발
2. 개발된 AI 모델을 CDSS에 적용하여 전공의 부담을 줄이고 숙련도 격차를 해소
3. 불필요한 건측 또는 다각도 x-ray 촬영 감소로 방사선 노출 최소화, 응급실 진료 효율 향상
4. 누구나 사용 가능하므로 지방 및 의료 취약 지역에서의 의료 서비스 개선

Keyword



방사선 노출 최소화  
신속·정확한 응급 진단 AI 기반



응급의료 체계에 AI기반 기술을 접목  
CDSS(Clinical Decision Support System) 의료 서비스 혁신

## 프로젝트 배경

# 연구동향

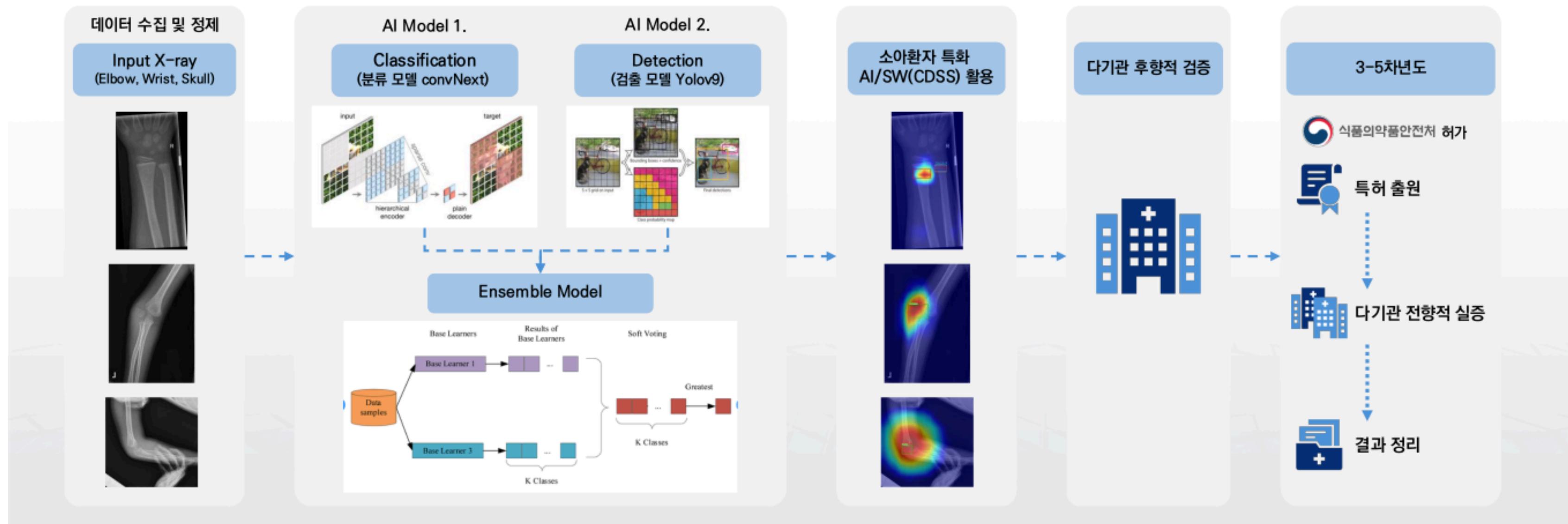
Paper	Network	Task	Testset	AUROC	ACC	F1	Sensitivity	Specificity
Using a Dual-Input Convolutional Network for Supracondylar Fracture Detection (SNUH/GNUCH) – Temporal	Dual-Input ResNet CNN	Binary (Supracondylar fracture)	Temporal Internal (same inst., 258 pairs)	<b>0.985 (0.962–0.996)</b>	-	-	93.9 (85.2–98.3)	92.2 (87.4–95.6)
Using a Dual-Input Convolutional Network for Supracondylar Fracture Detection (SNUH/GNUCH) – Geographic	Dual-Input ResNet CNN	Binary (Supracondylar fracture)	Geographic external (different hospitals)	0.992 (0.947–1.000)	-	-	100.0 (85.2–100.0)	86.1 (75.9–93.1)
Detection of Traumatic Pediatric Elbow Effusion Using a Deep Learning Model	VGG-16	Effusion detection (Fat Pad)	5-fold CV	0.943 (0.884–1.000)	0.908 (0.84–0.95)	0.91	0.909 (0.82–0.97)	0.906 (0.83–0.96)
Binomial Classification of Pediatric Elbow Fractures Using a Deep Learning Model	CNN-Xception, RNN	binomially classified, based on the training dataset, of which 4966 negative		0.9465	88.0(86.0, 90.0)	-	91.0(88.0 , 93.0)	84.0(81.0 , 87.0)
External validation of a commercially available mobile application for fracture detection in children	Rayvolve		Elbow (nsets= 294)		89%		92%	87%
Knowledge-Guided Multiview Deep Learning for Fracture Detection	Curriculum CNN	Fracture classification	Internal split	0.974 (-)	0.938 (-)	0.94	0.941 (-)	0.932 (-)
Fractured Elbow Classification Using a Deep Learning Model	ResNet-169 / InceptionV3 / YOLOv5 / U-Net	Review (Cls/Det/Seg)	Internal validation (MURA)	0.96	97.1	97	97	96
Automated detection of acute appendicitis using deep learning	BoneView (Gleamer) , Detectron2		(hand/wrist, elbow/upper arm, shoulder)	0.93			91.3% [85.6, 95.3]	90.0% [84.0,94.3]
Deep learning accurately classifies elbow fractures	VGG-16 + DCNN	Effusion detection (Fat Pad)	Internal split(8:2)	0.951 (0.946–0.955)	0.900 (-)	-	0.930 (0.91–0.95)	0.870 (0.85–0.89)
Use of deep learning model for paediatric elbow fractures	EfficientNet B1	Prognostic (stiffness)	test set (112 images)	0.872 (0.831–0.947)	80.4 (71.8–87.3)	-	79.0 (68.40–89.5)	81.8 (71.6–92.0)
Construction and validation of elbow fracture detection model after supracondylar humerus fracture	Logistic regression analysis	Binary (Fracture presence)	Internal Validation (7:3 )	0.8035 (95% CI:0.79–0.93)	-	-	-	-
Artificial intelligence to identify fractures in children	R-CNN /EfficientNetV2-Small		external test (n = 515) splits	<b>0.96 (95% CI 0.95–0.97)</b>	89.7% (95% CI 88.0–91.3%)		90.8% (95% CI 88.5–93.1%)	88.7% (95% CI 86.4–91.0%)
Comparison of diagnostic performance of deep learning models for elbow fractures	VGGNET		-	0.93(0.92-0.94)			0.957	0.916(0.89)
Meta-analysis: DL for Pediatric Elbow Fracture Screening	DL for Pediatric ElbowNet/VGG/Detectron2/Xception (various)	Fracture screening (pooled)	6-study pooled internal/external	0.95	-	-	0.93	0.89

# **프로젝트 개요**

# 주요기능

## 소아청소년 X-ray 골절 판독 모델 개발 진행 상황

기존 보고된 연구 성능보다 뛰어난 판독 모델을 개발하기 위해 BBox / heat map 을 추가하여 explainability 를 확보했고 각각의 성능을 서로 보완하기 위해 ensemble 을 진행.  
이 두 작업 모두 소아 영역에서는 케이스가 적고, 병변사진은 더 희귀하기 때문에 특장점으로 작용함.



# 기술 사항



# 기술

## Pre-processing

- Dataset
- Image Pre-process
- Workflow Pipeline

1

## Classification Model

- Swin-T
- ConvNeXt
- Grad-CAM Results

2

## Detection Model

3

- YOLOv9
- Overlay Results

## Ensemble Model

4

- Soft Voting system
- Final Results

# Pre-processing

## 1. Dataset

- Wrist
  - MURA(성인) : 9,752 case (Fracture : 3,987 , Normal : 5,765)
    - 골절 여부만 존재
  - GRAZPEDWRI-DX(소아) : 20,326 case (Fracture : 12,957 , Normal : 6,661)
    - AO + Age + Gender 등 풍부한 메타data
  - 병원(성인, 고려대임상셋) : 2,100 case (Fracture : 683 , Normal : 1,417)
    - 부위별 cropped 이미지 (radius, scaphoid, styloid)
- Elbow
  - 병원(소아, 아산병원) : 6,758 case (Fracture : 2,203 , Normal : 4,555)
    - AP : 3,324 case (Fracture : 1,042 , Normal : 2,282)
    - LAT : 3,434 case (Fracture : 1,161 , Normal : 2,273)

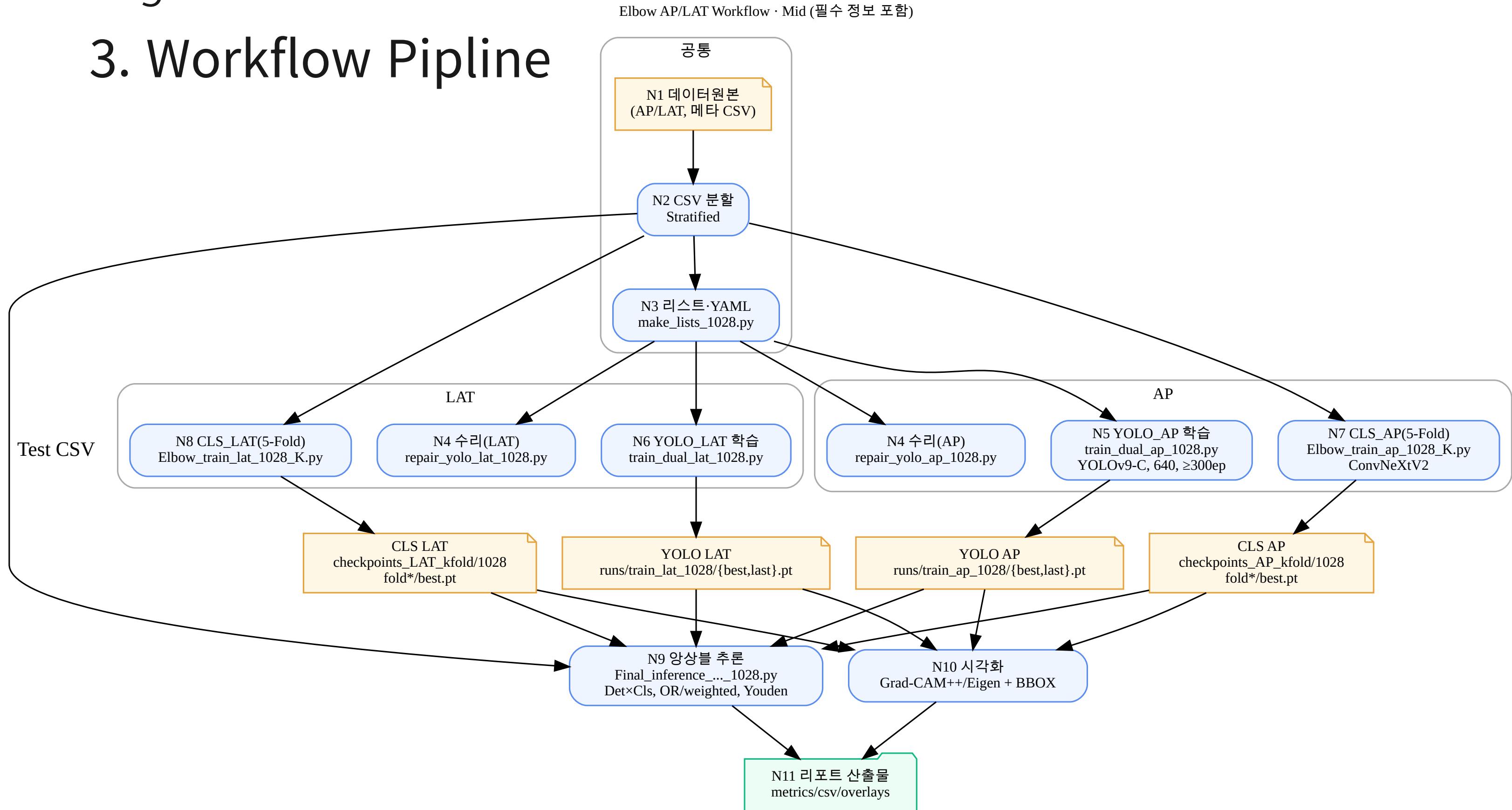
# Pre-processing

## 2. Image Pre-process

- Image resize
  - ConvNext : 384 x 384 x 3 (Wrist) , 512 x 512 x 3 (Elbow)
  - YOLO : 512 x 512 x 3 (Wrist, Elbow)
- Transformation
  - Default : hsv, brightness, CLAHE 등등
  - Flipping : Elbow
  - Cropping : 추후예정
- Input Data Split
  - 환자단위 Test set 우선 분류
  - Train / Val = 8 : 2 분리
  - AP\_train.csv, LAT\_train.csv , AP\_val.csv 등등 저장
  - image\_path, Labels, Fracture 등등 메타 데이터 같이 저장

# Pre-processing

## 3. Workflow Pipeline

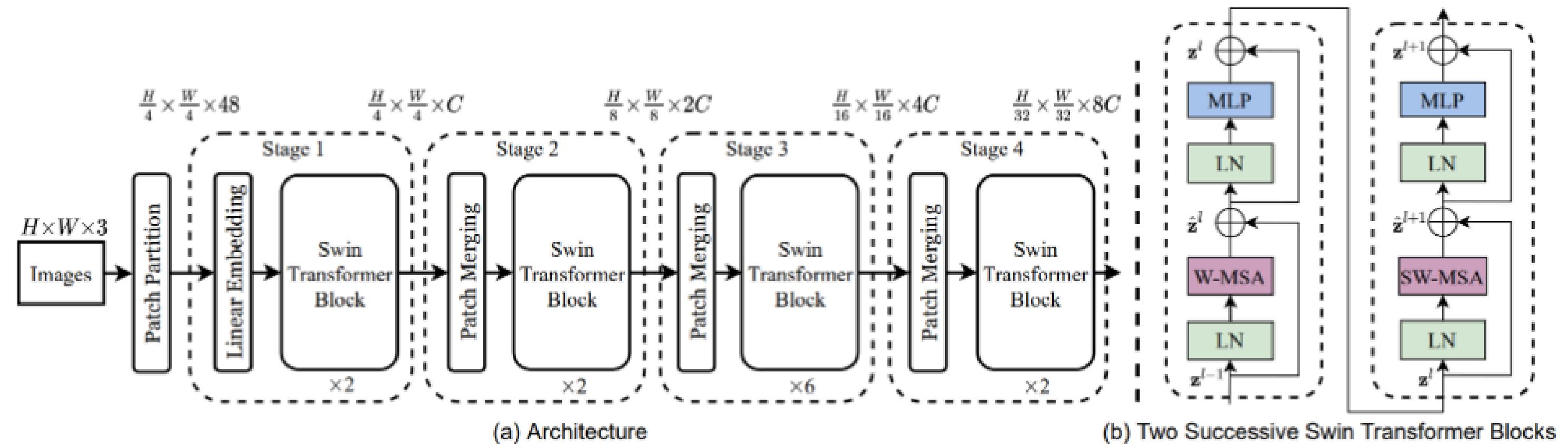


## Classification Model

## 1. Swin-T

## • Architecture

- $4 \times 4$  패치 → 토큰.
- 4단계 피라미드: W-MSA  $\leftrightarrow$  SW-MSA 반복, 각 단계마다 Patch Merging으로 해상도  $\frac{1}{2}$ , 채널  $2 \times$ .
- Train 정확도 84.45% Val 73.36%, Specificity = 0.6075, Sensitivity = 0.8334
- Test 정확도 68.33%, Specificity = 0.4833, Sensitivity = 0.8833

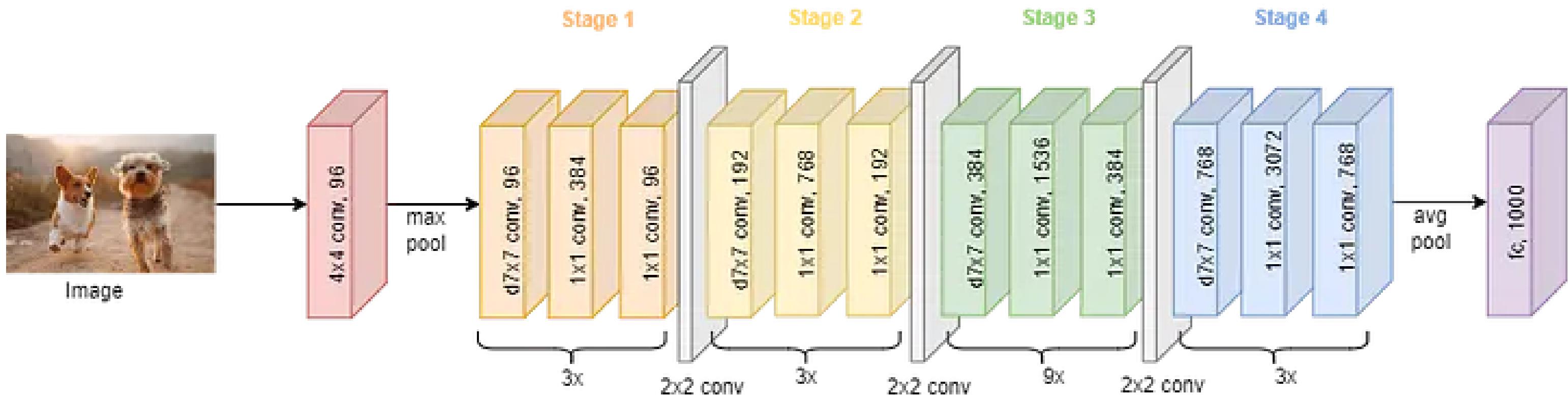


## Classification Model

## 2. ConvNeXt

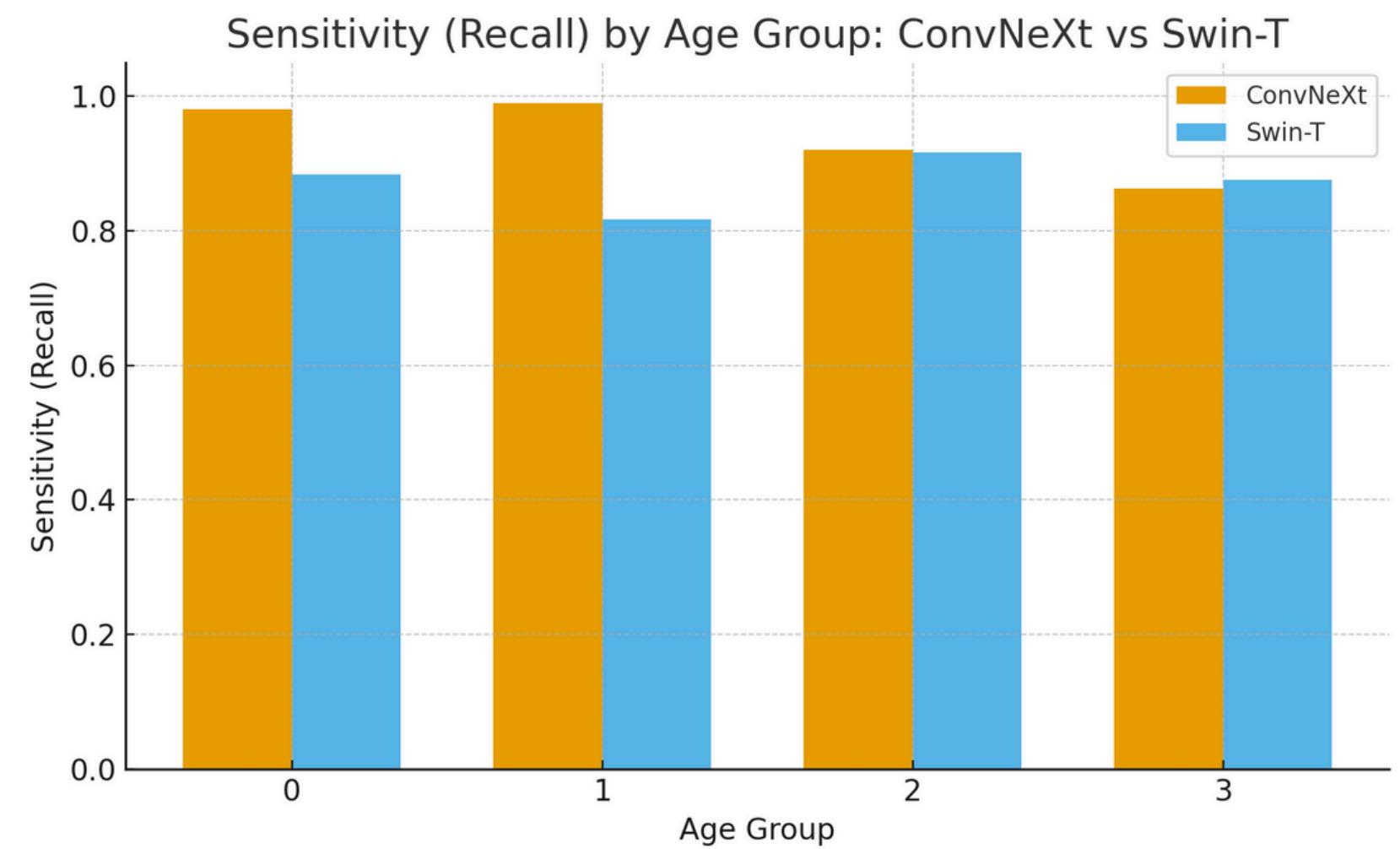
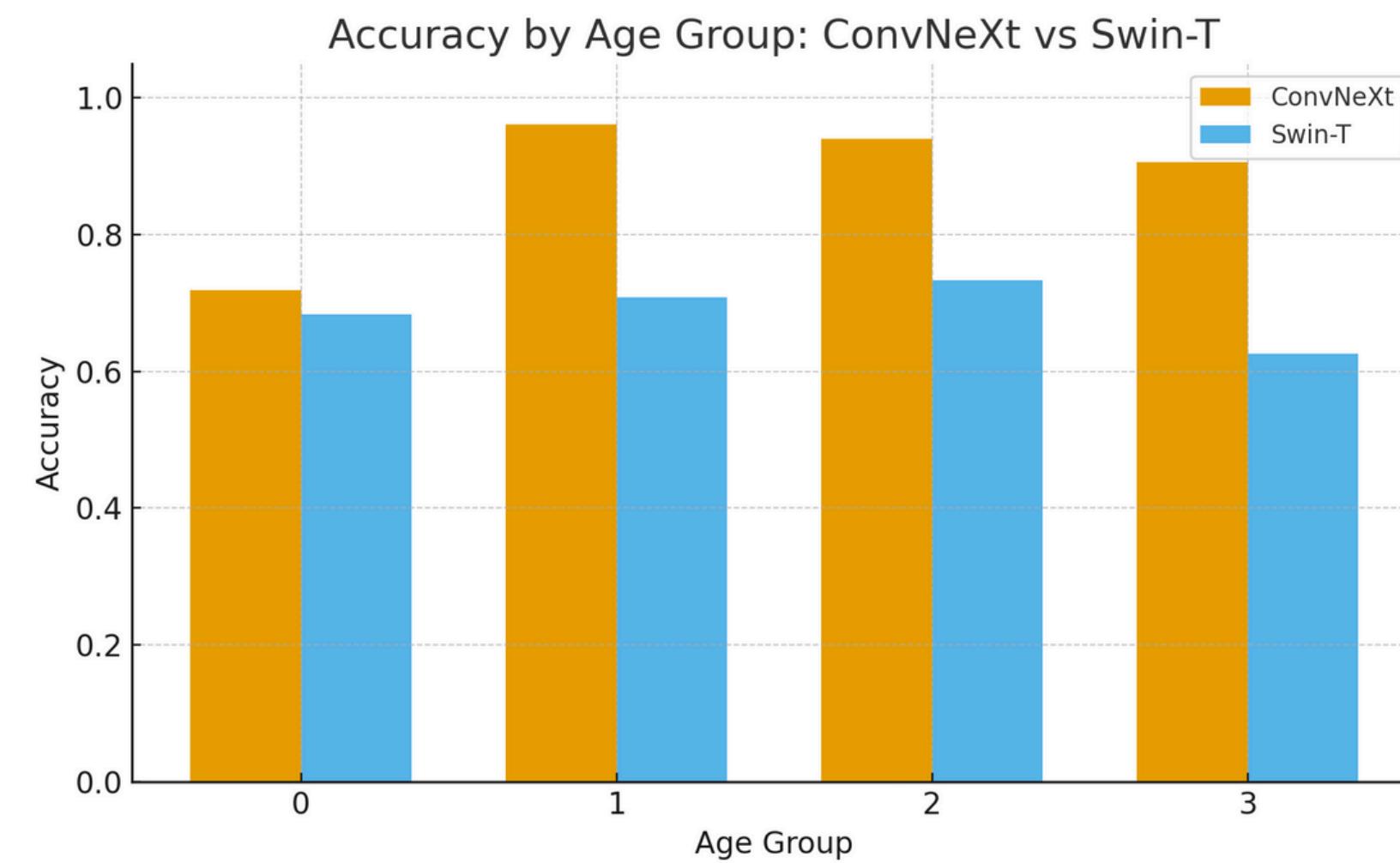
## • Architecture

- 큰 커널( $7 \times 7$ ) 심층 컨볼루션 + Depthwise/Pointwise 분리.
- LayerNorm 사용, GELU, 마이크로 디자인 단순화
- Train 98.8%, Val 90.3%, Sens 93.8%, Spec 82.1%



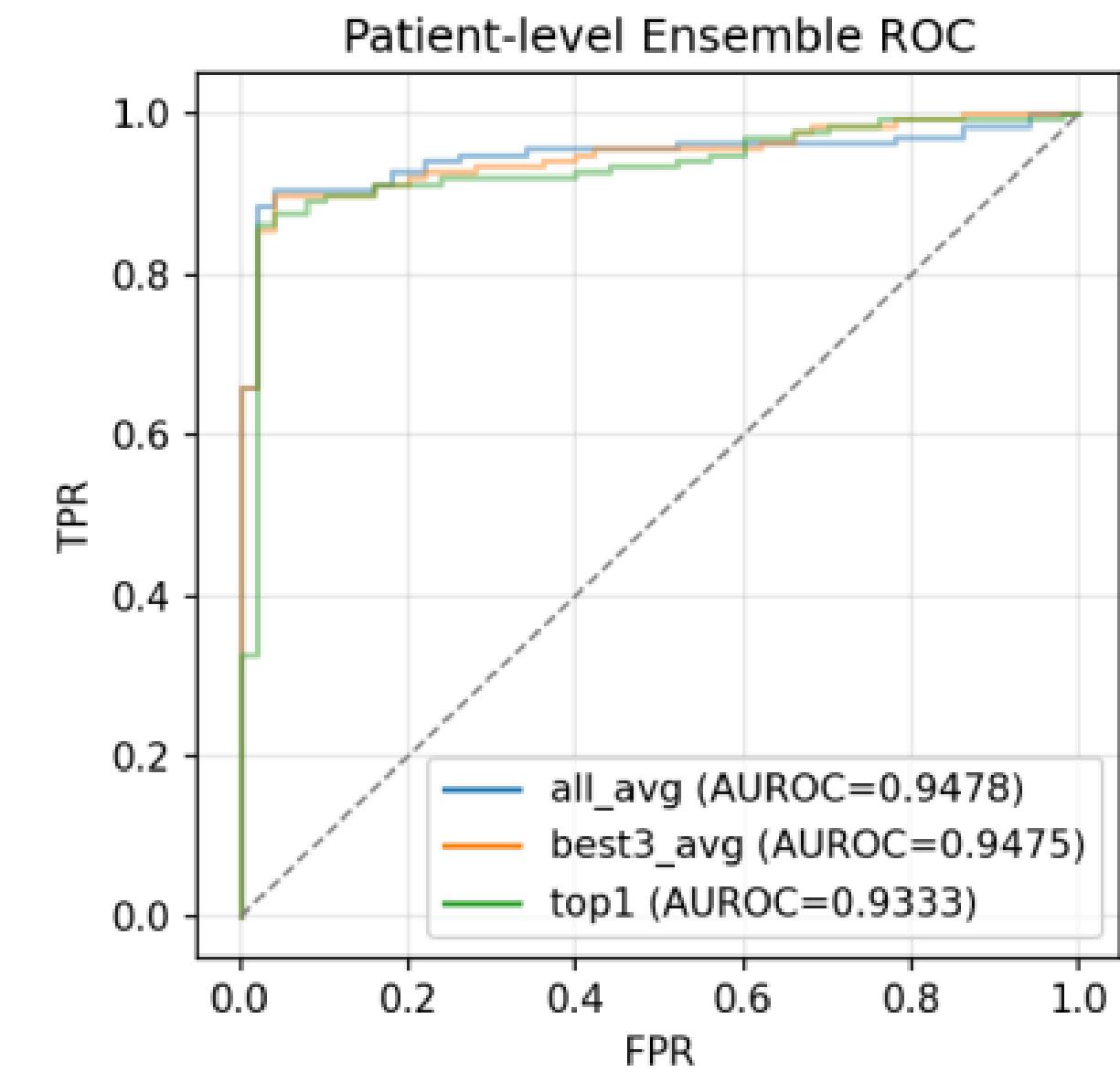
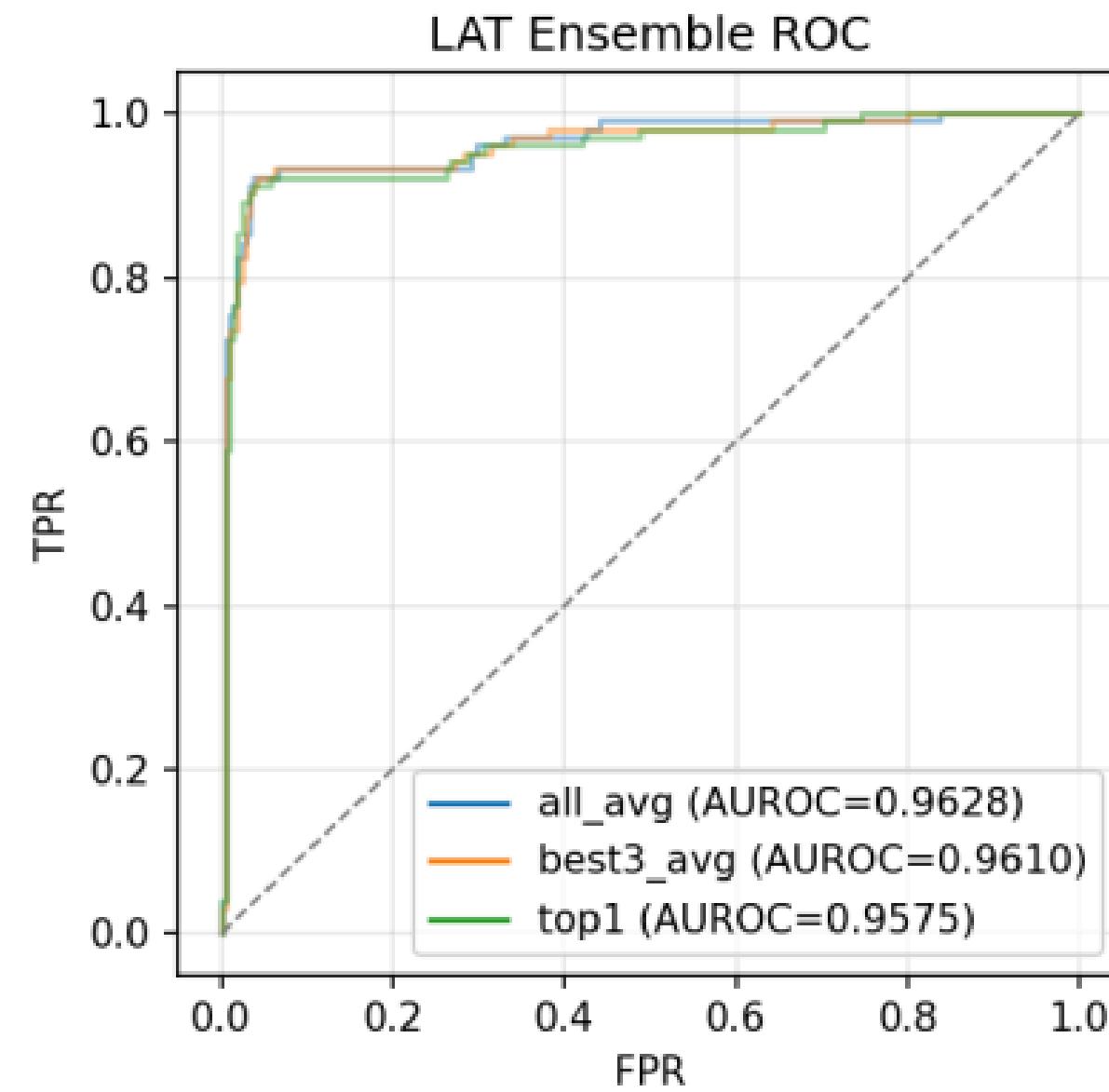
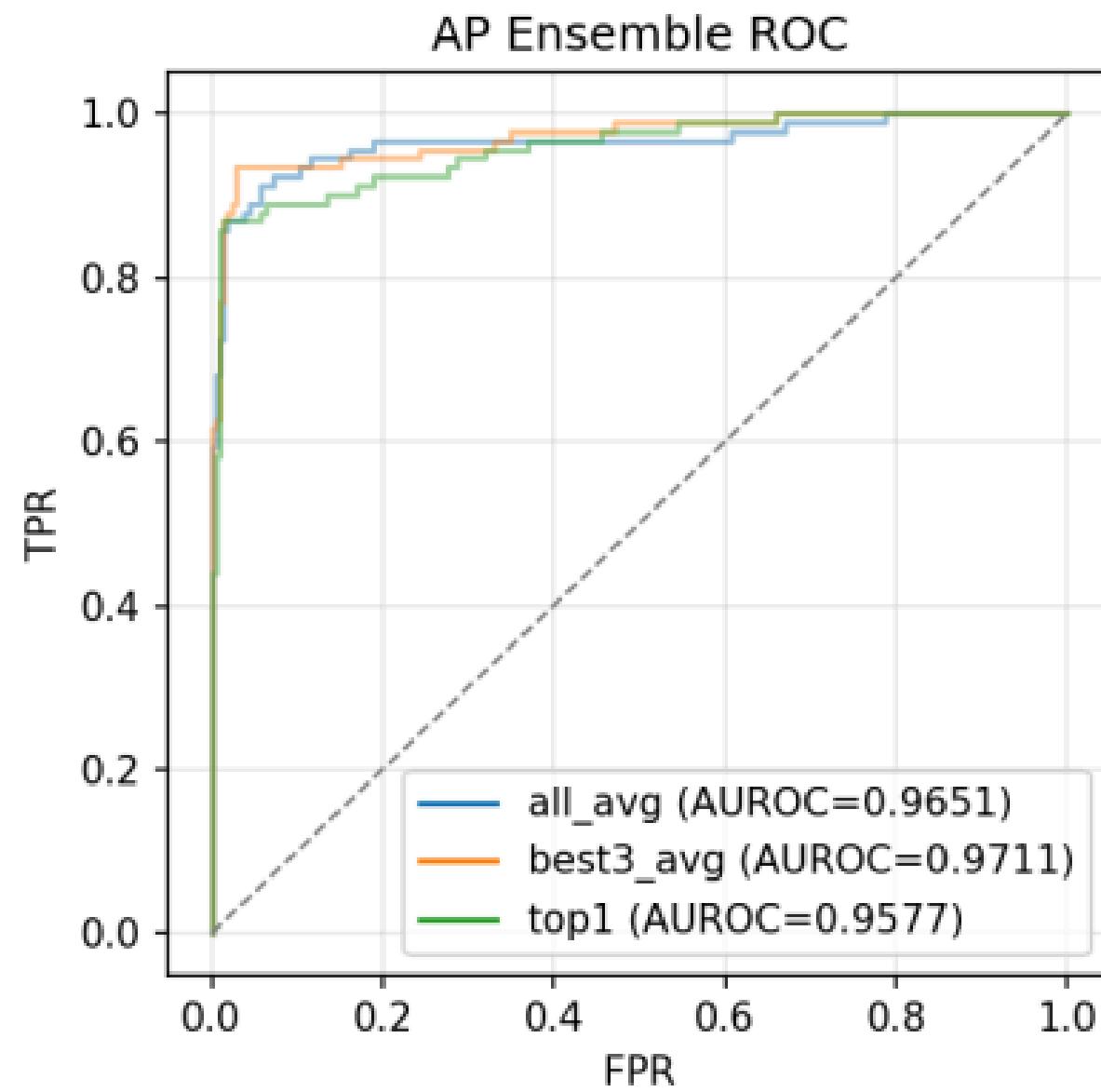
## Classification Model

## 3. Results

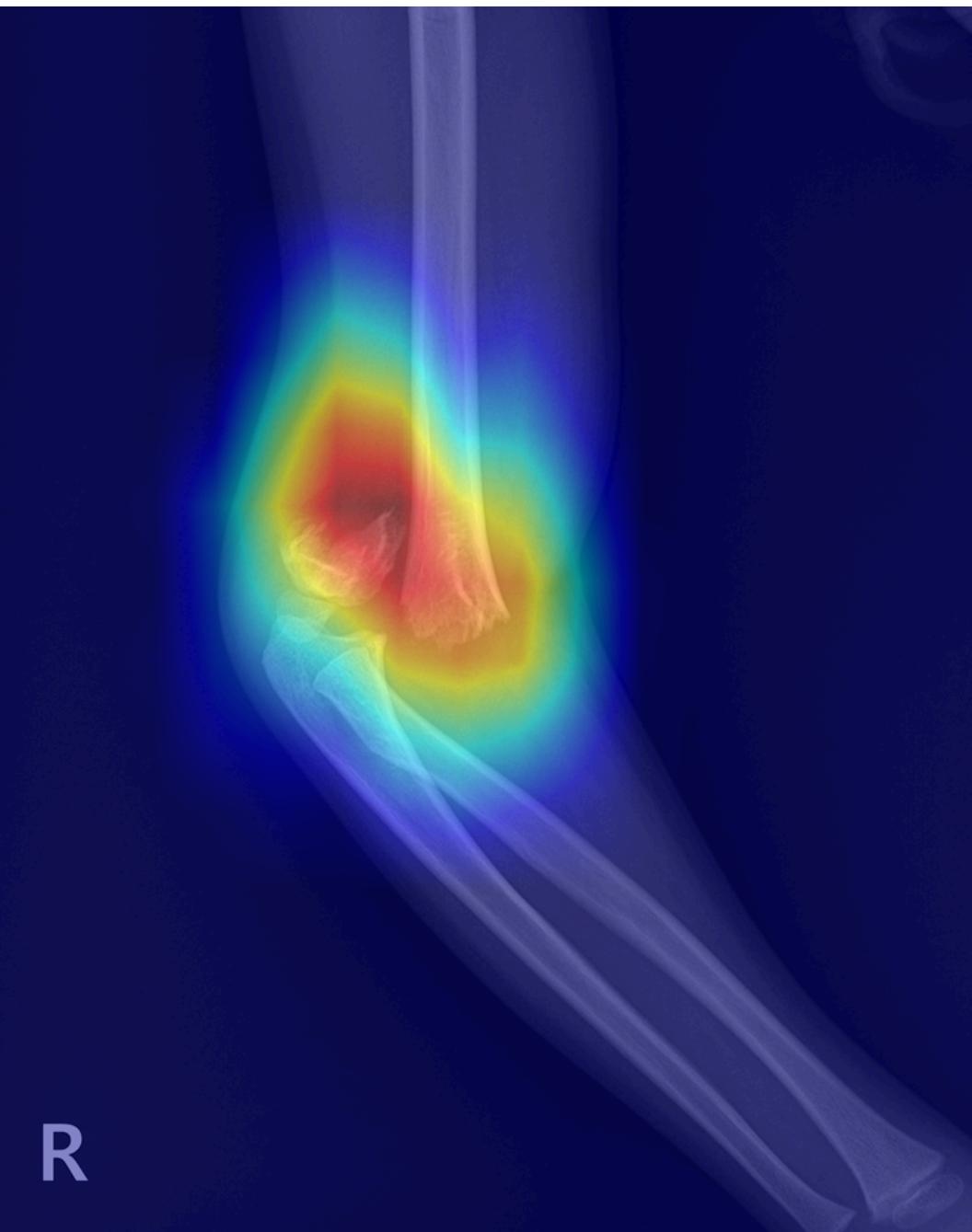


## Ensemble Model

## 3. Result (5-Fold Cross Validation)



### 3. Results



# Classification Model

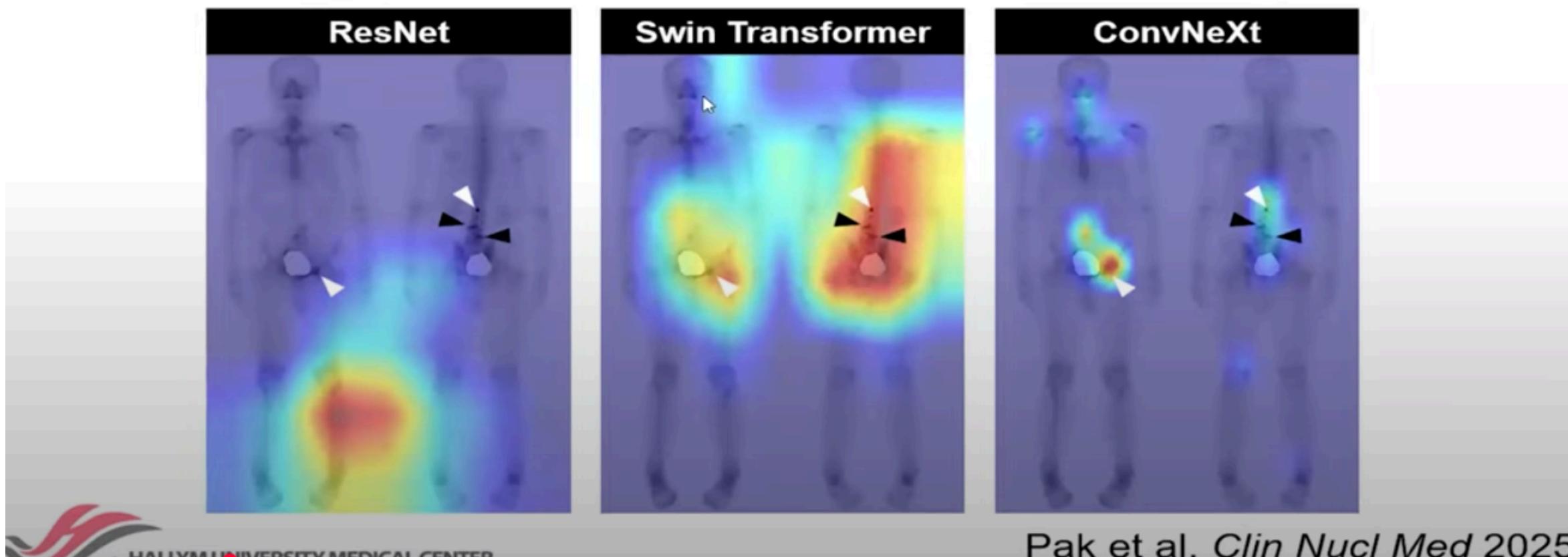
## recommend

### Comparison of CNNs and Transformer Models in Diagnosing Bone Metastases in Bone Scans Using Grad-CAM

Sehyun Pak,\* Hye Joo Son, MD, PhD,† Dongwoo Kim, MD,‡

Ji Young Woo, MD, PhD,§ Ik Yang, MD, PhD,§

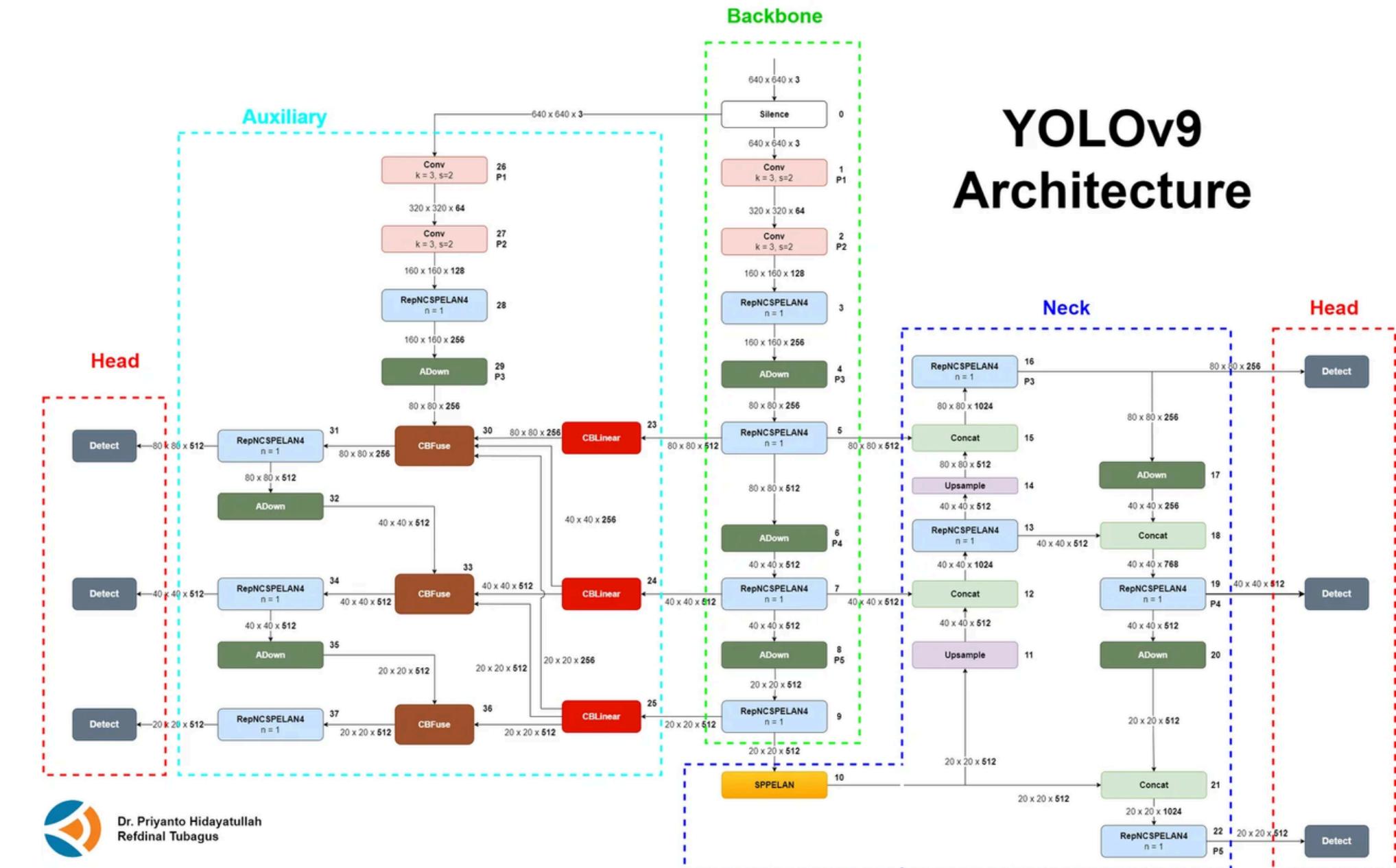
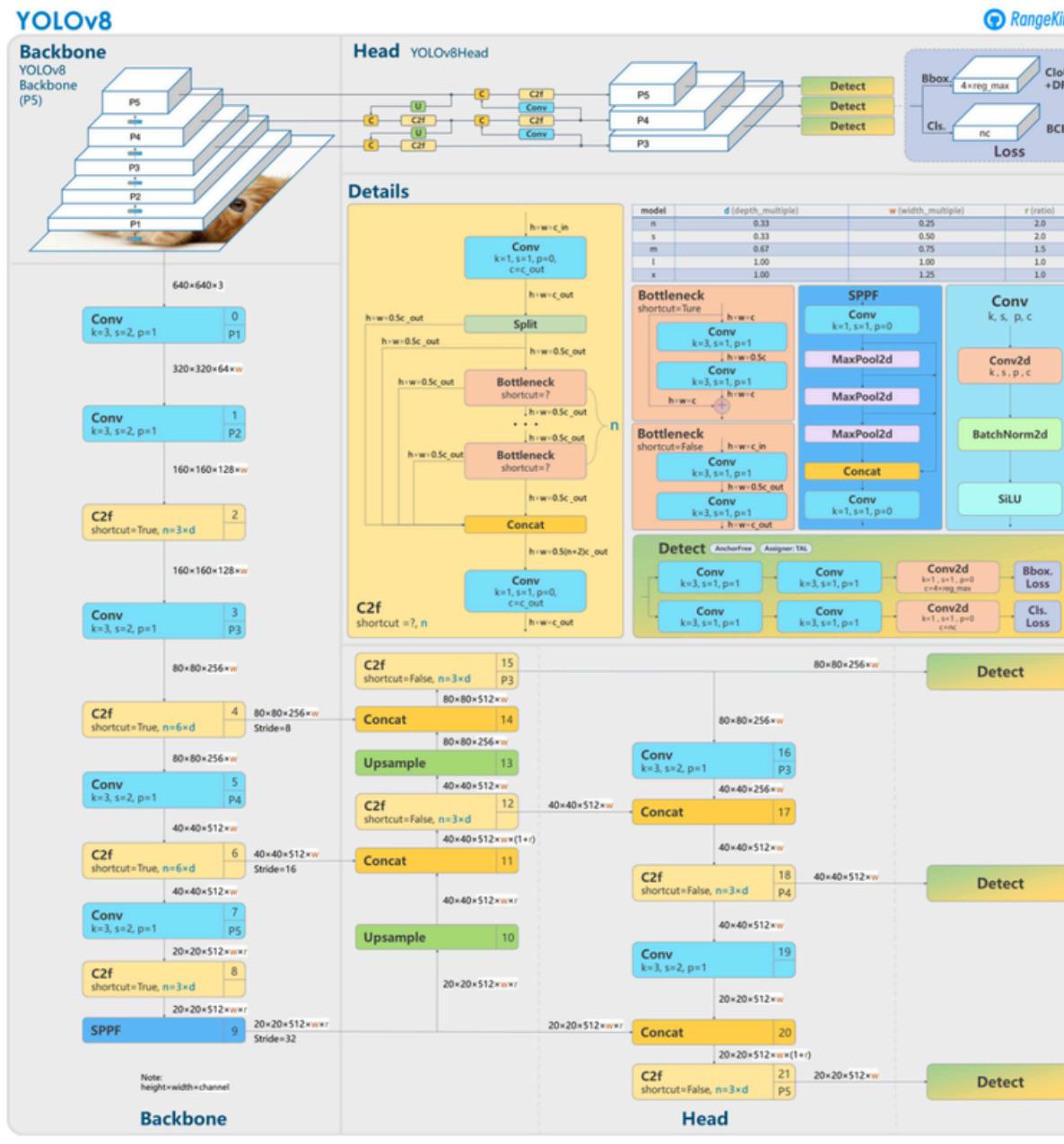
Hee Sung Hwang, MD, PhD,‡ Dohyoun Rim, MS,|| Min Seok Choi, BS,¶ and  
Suk Hyun Lee, MD, PhD§



<https://www.ibric.org/bric/webinar/academic-all.do?mode=view&id=940>

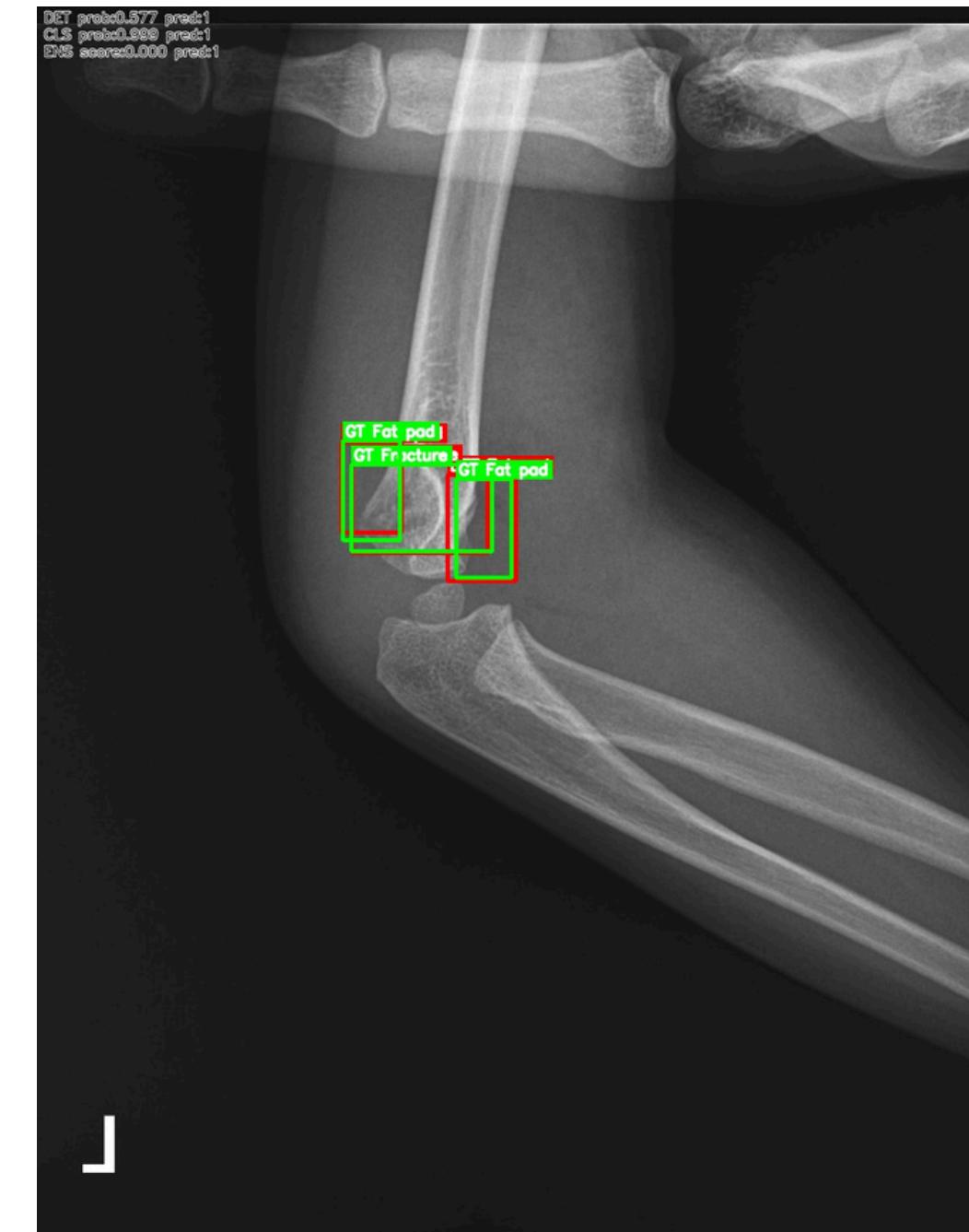
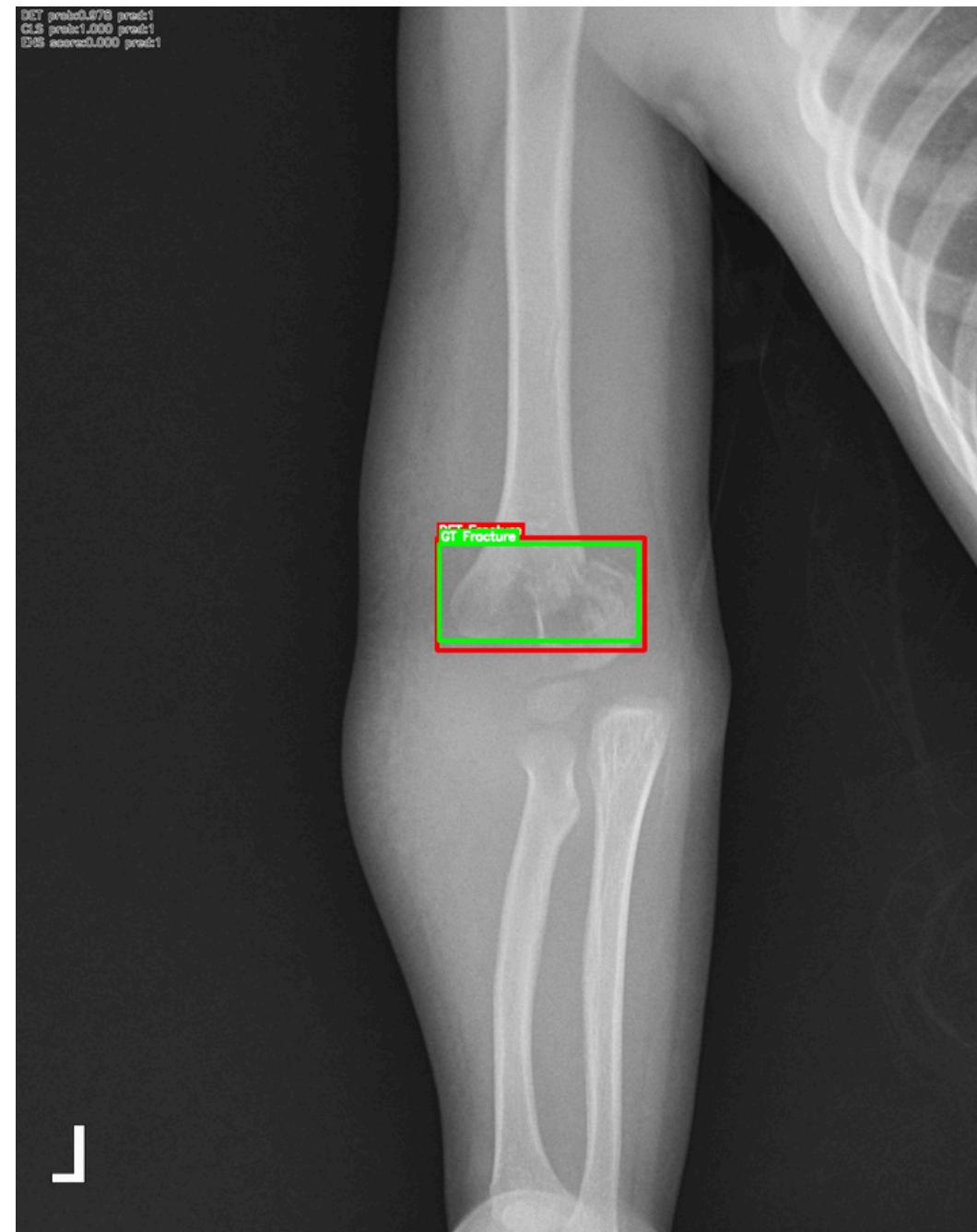
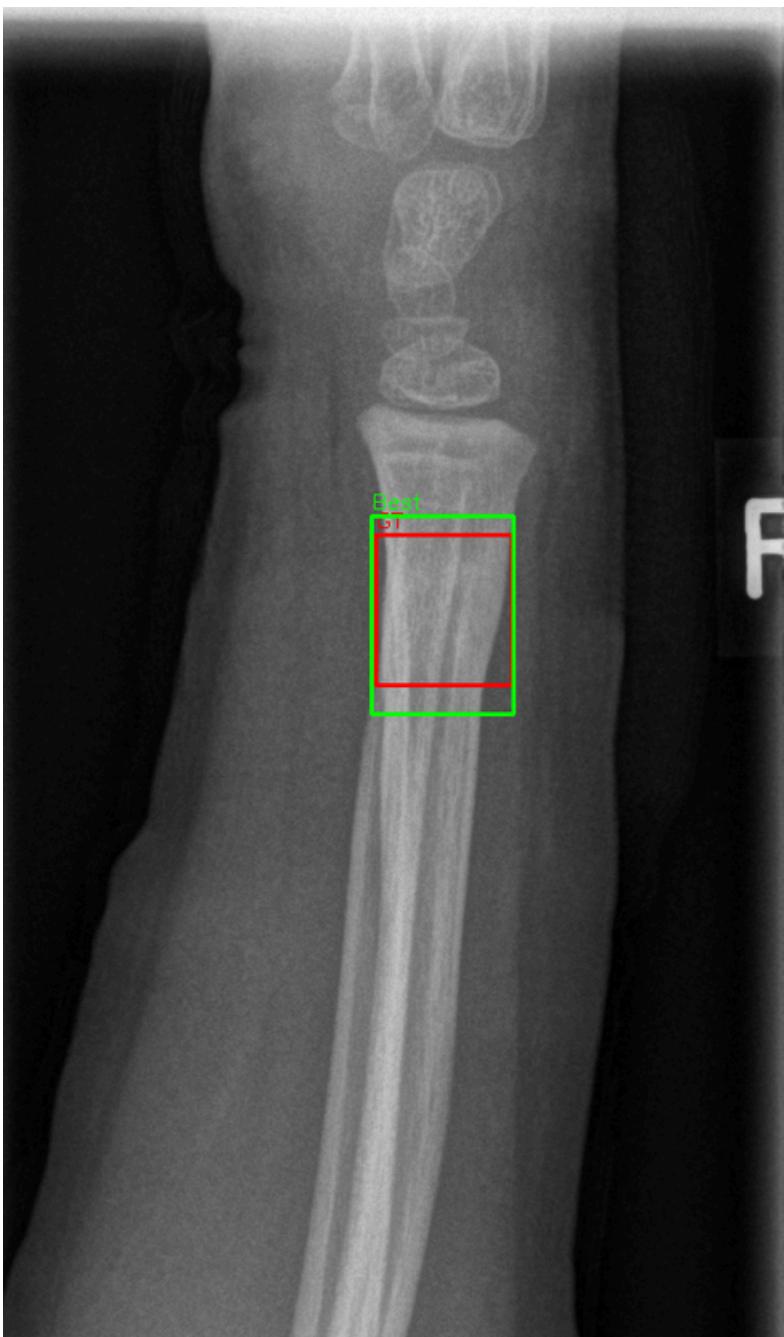
# Detection Model

## 1. YOLO



# Detection Model

## 2. Result



---

# Ensemble Model

## 1. Voting System

### 2) Hard voting

- Unanimous voting : 만장일치일 때만 선택한다.
- Majority voting : 제일 많이 나온 것을 선택한다.

$$\text{Ensemble}(\hat{y}) = \operatorname{argmax}\left(\sum_{j=1}^n I(\hat{y}_j = i), i \in (0, 1)\right)$$

### 3) Soft voting

- 각 class별로 모델들이 예측한 확률의 평균 값을 계산하여 가장 높은 class 선택

$$\text{Ensemble}(\hat{y}) = \operatorname{argmax}\left(\frac{1}{n} \sum_{j=1}^n P(y = i), i \in (0, 1)\right)$$

### 4) Weighted voting

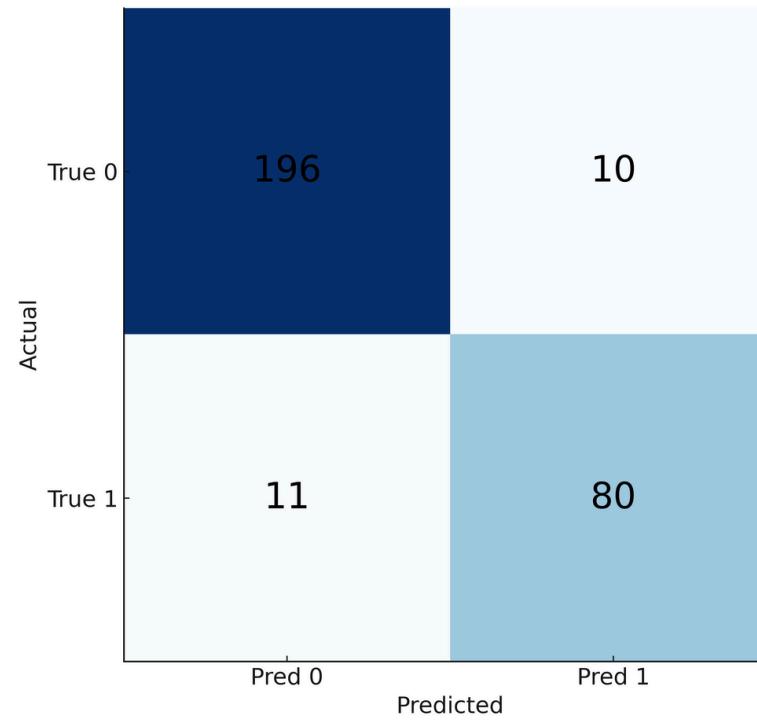
- 각각의 모델별로 class에 대한 가중치를 준다.
- 아래 식은 training accuracy를 사용하여 가중평균을 계산한 방법이다.

$$\text{Ensemble}(\hat{y}) = \operatorname{argmax}\left(\frac{\sum_{j=1}^n (\text{TrainAcc}_j) \cdot I(\hat{y}_j = i)}{\sum_{j=1}^n (\text{TrainAcc}_j)}, i \in (0, 1)\right)$$

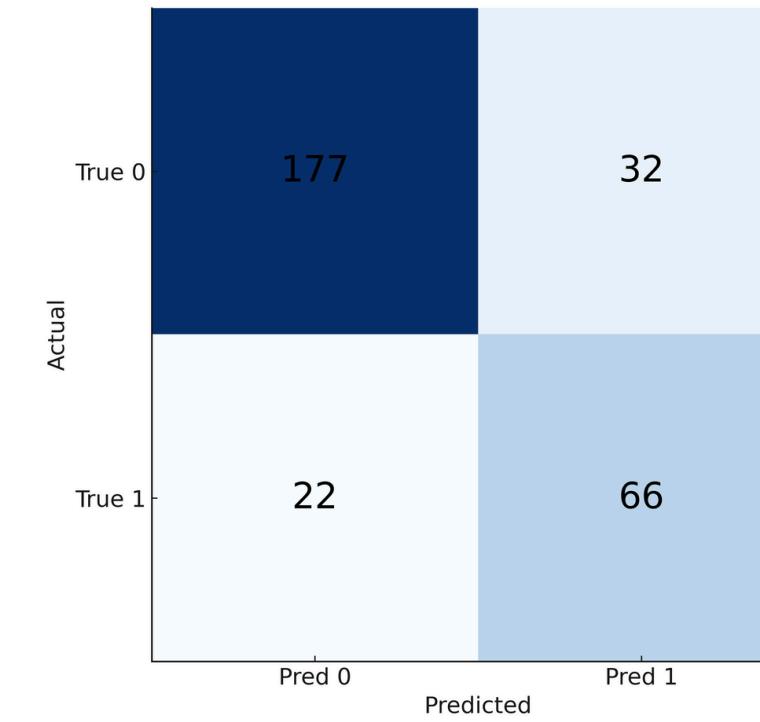
# Ensemble Model

## 2. Result

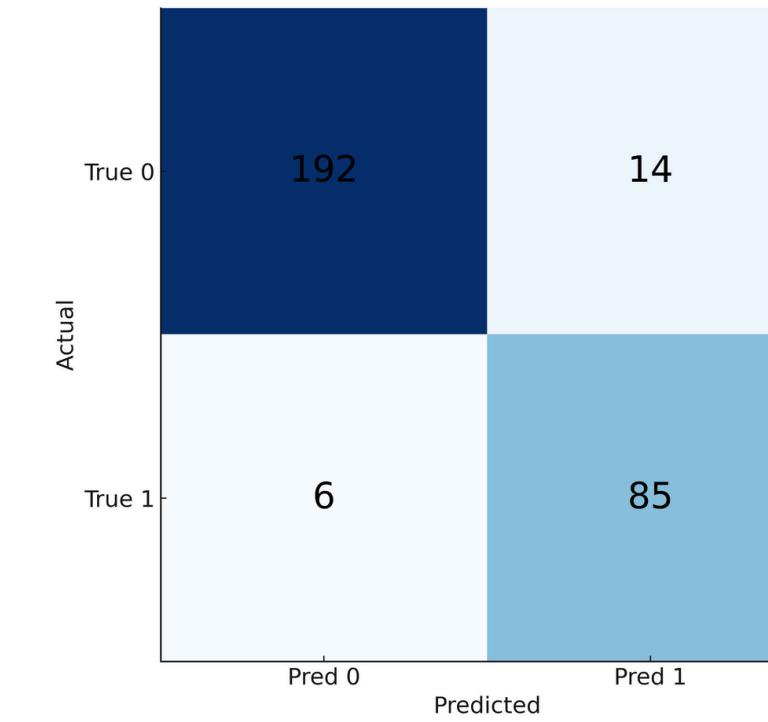
AP ConvNeXt



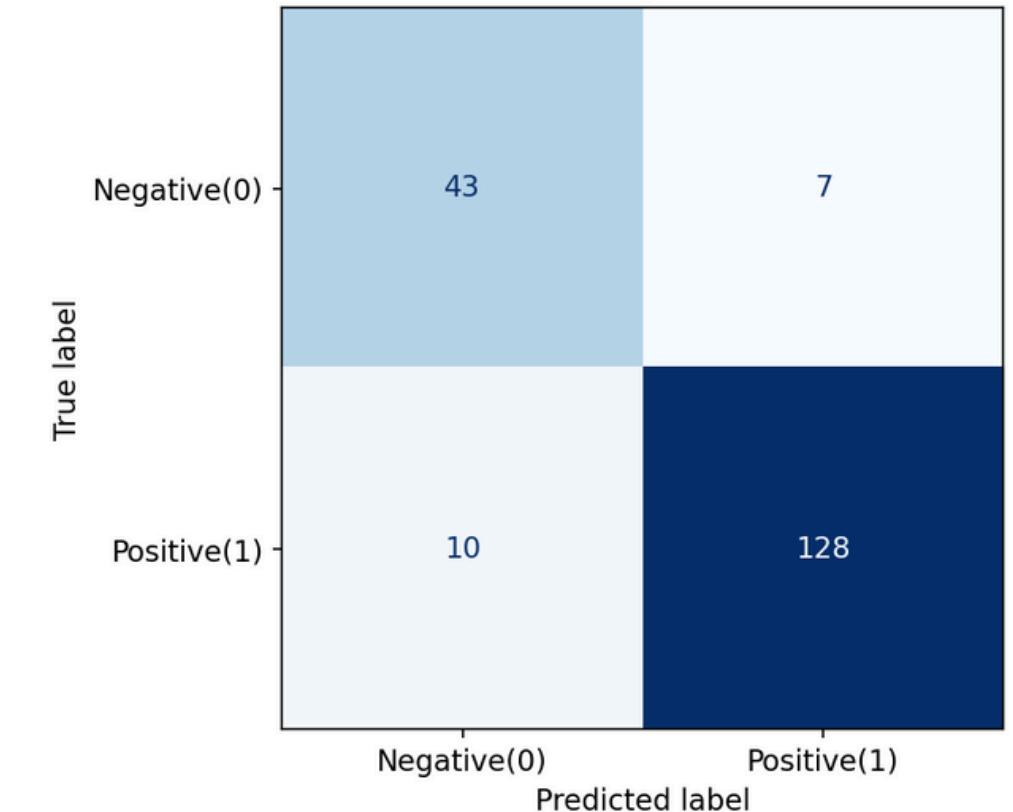
AP YOLO



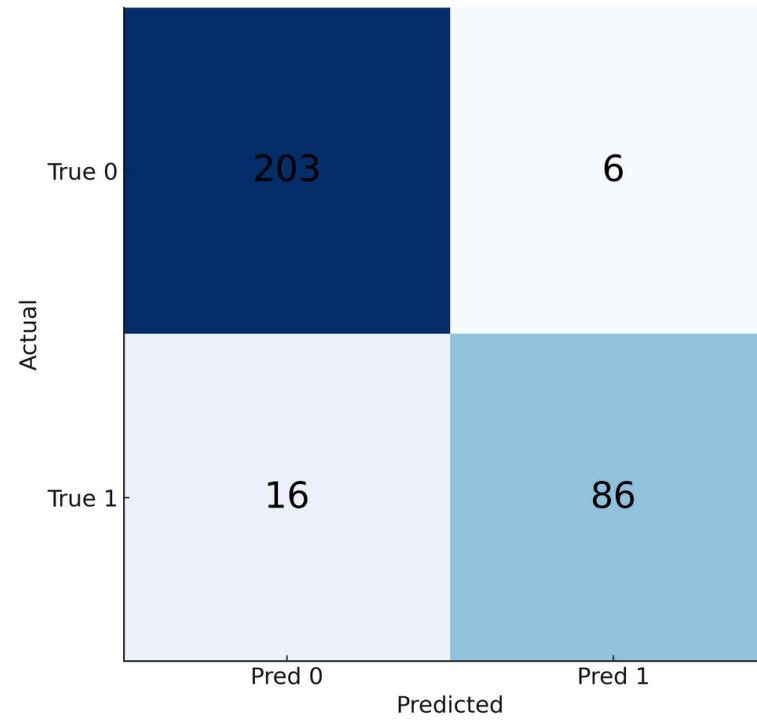
AP Ensemble



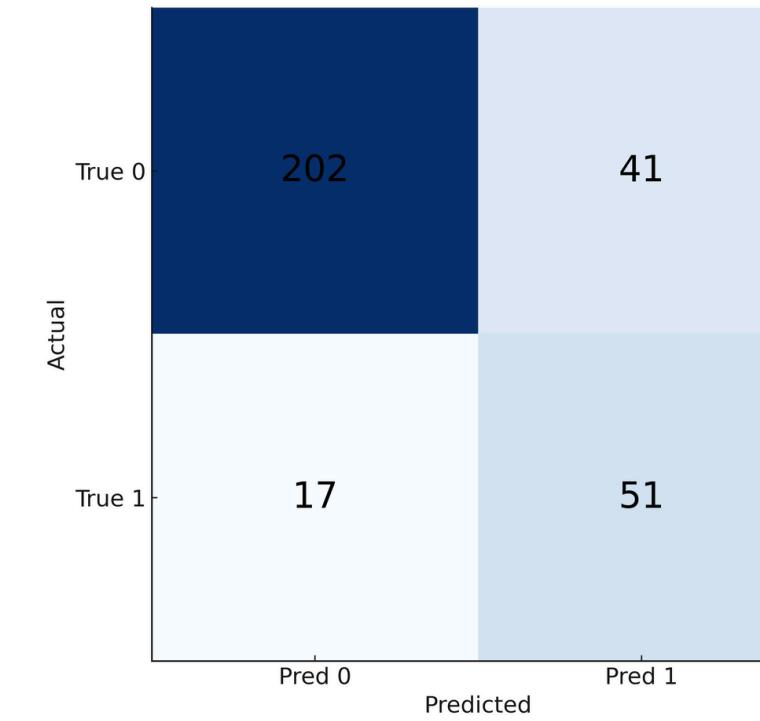
Patient-level Ensemble (ANY AP/LAT → Fracture)



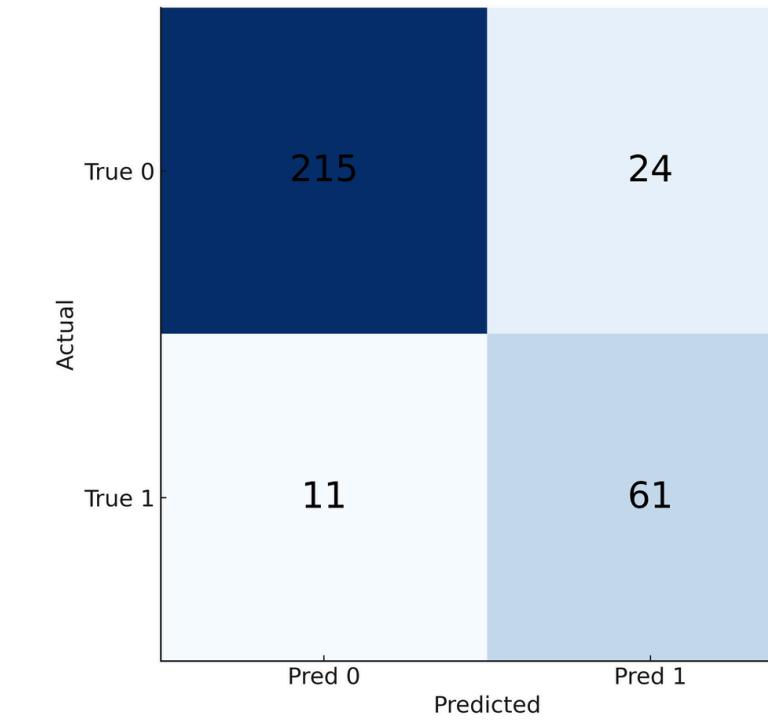
LAT ConvNeXt



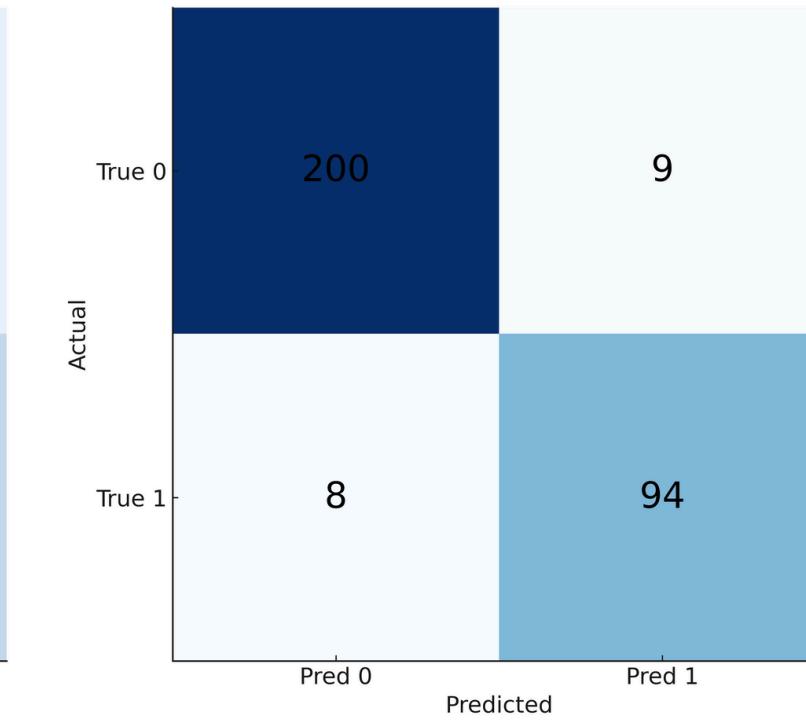
LAT YOLO (Fracture)



LAT YOLO (Fat pad)

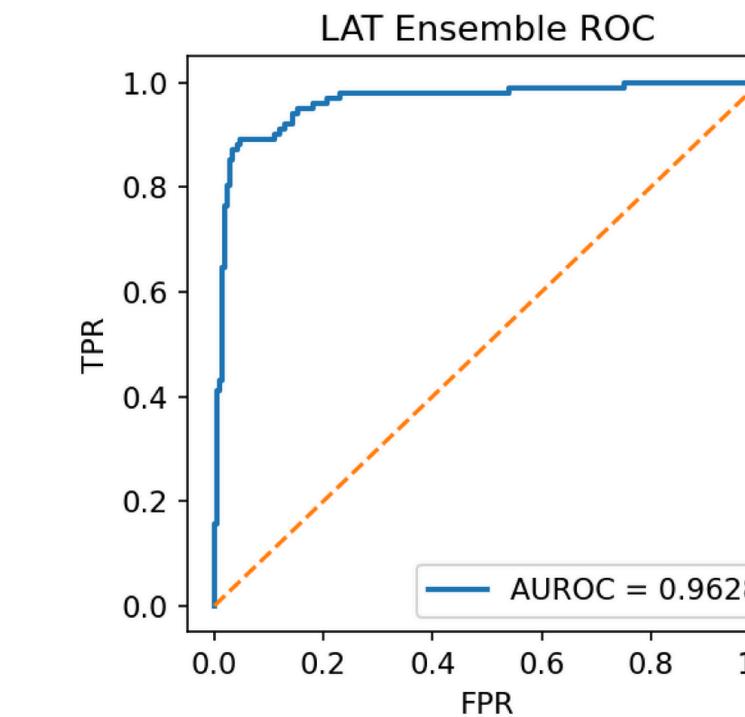
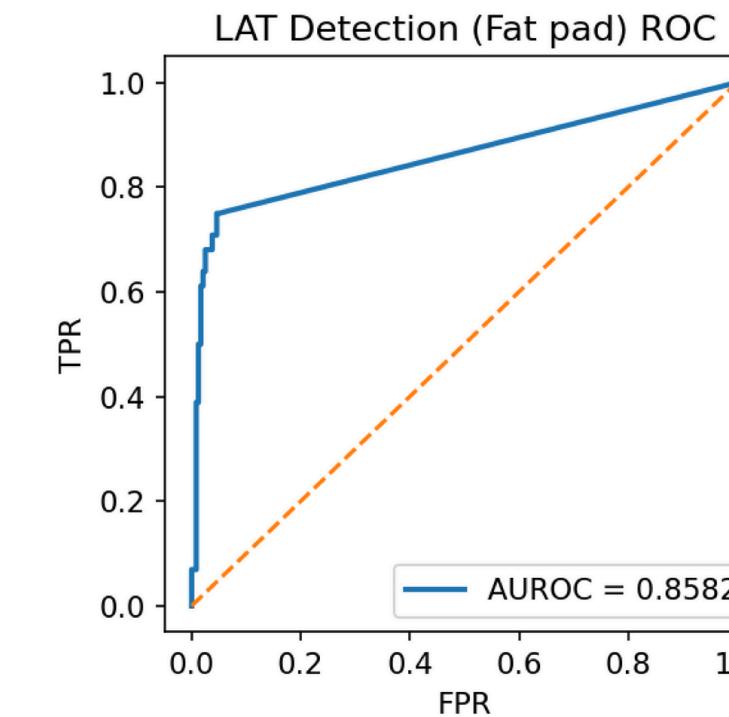
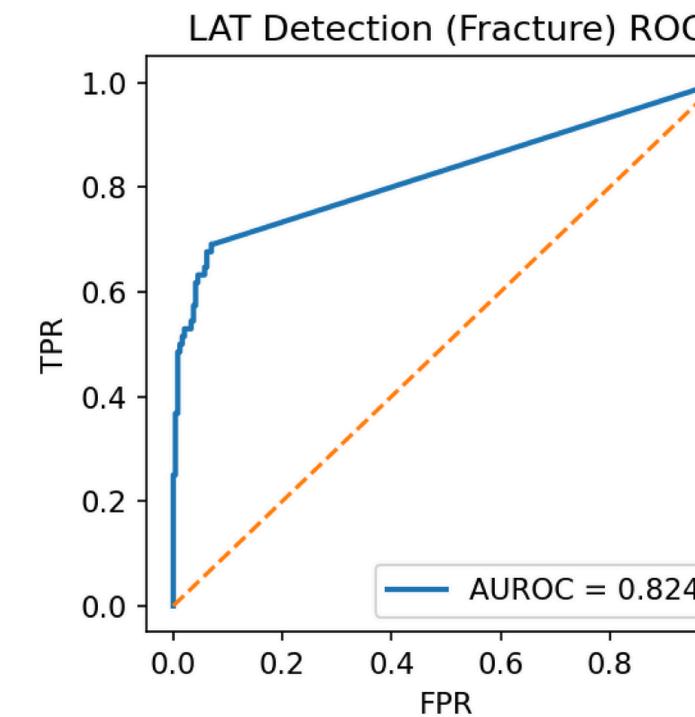
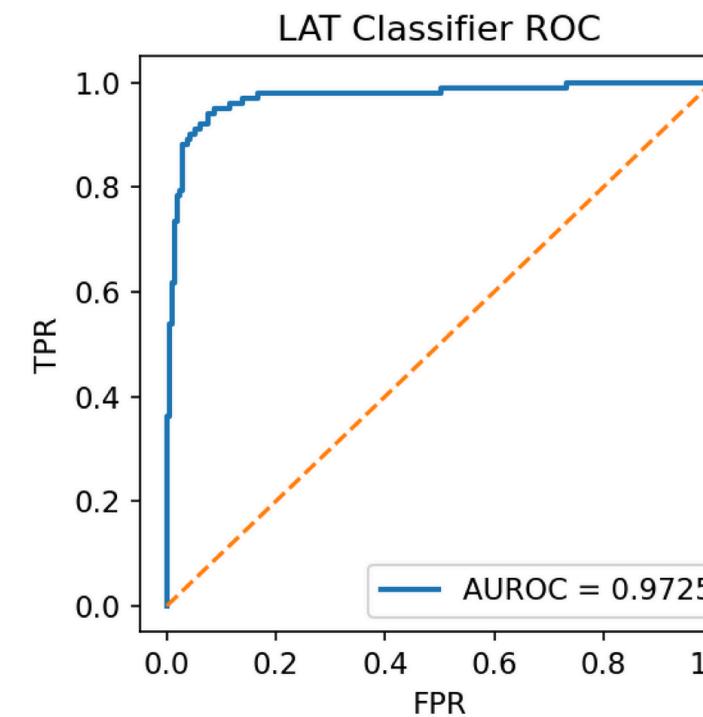
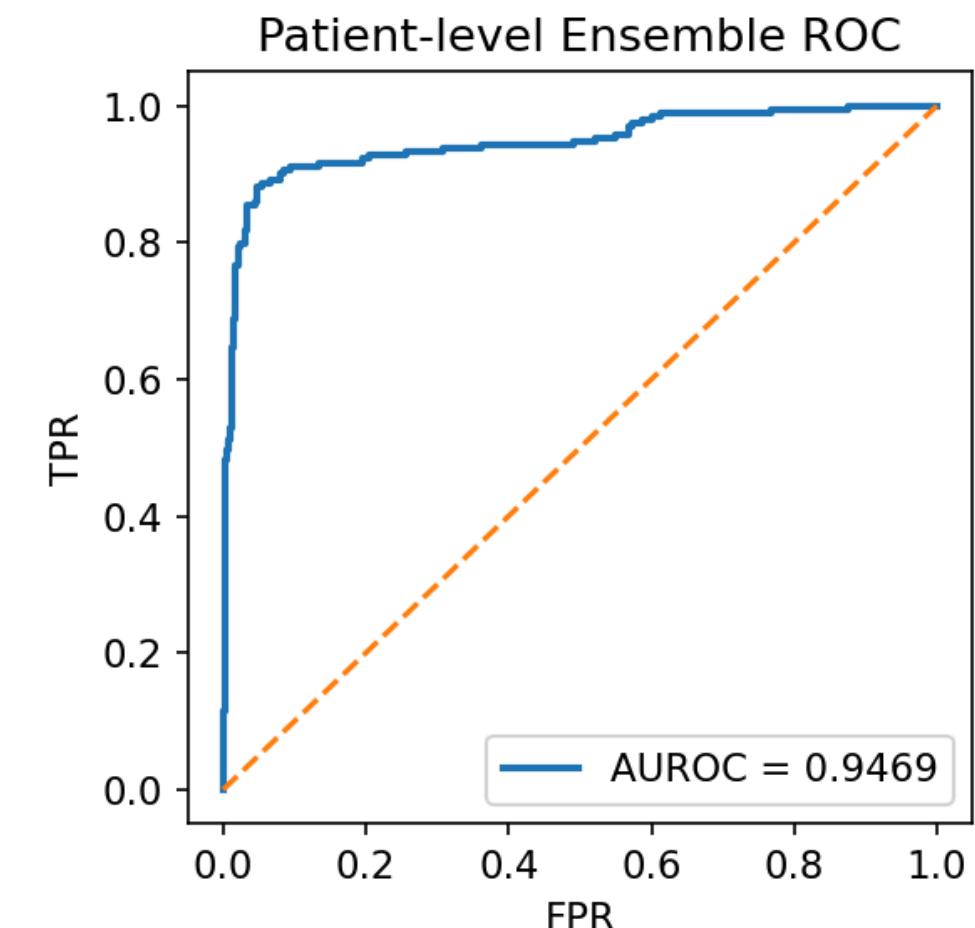
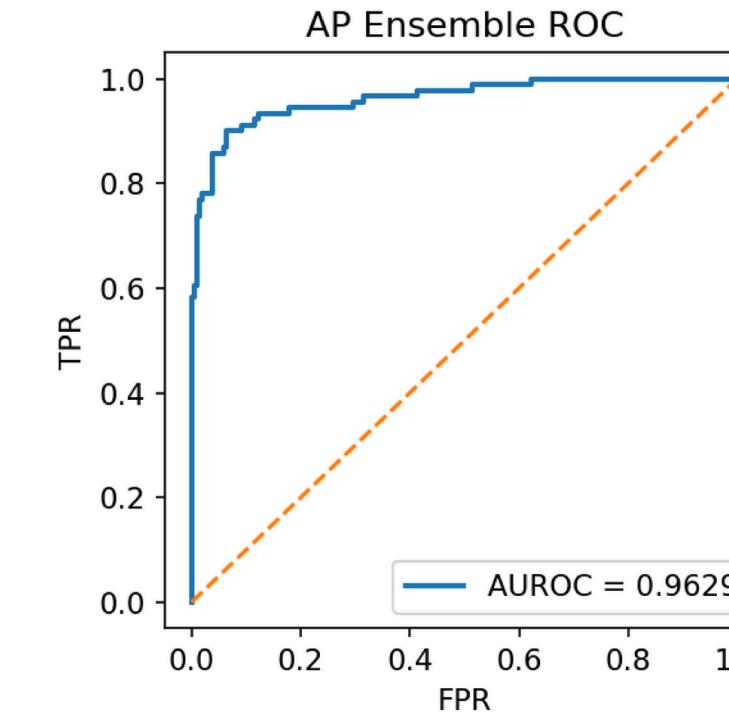
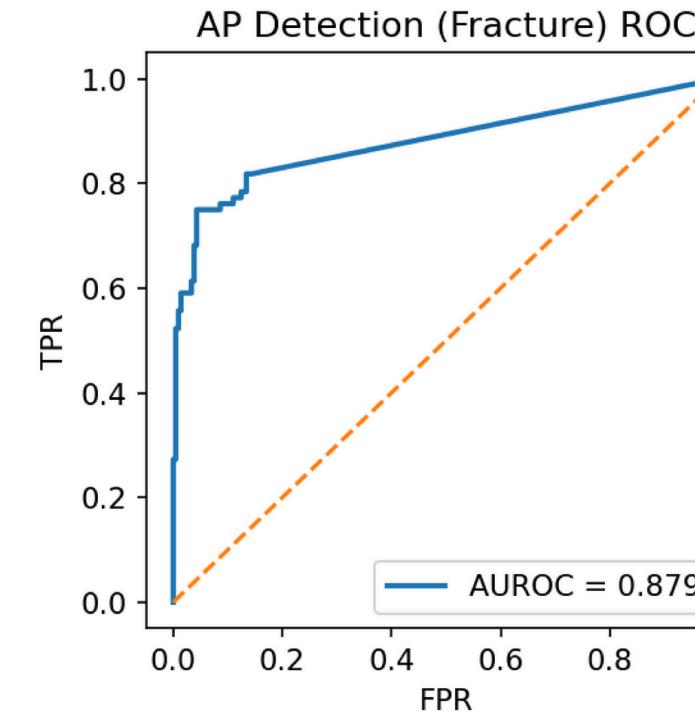
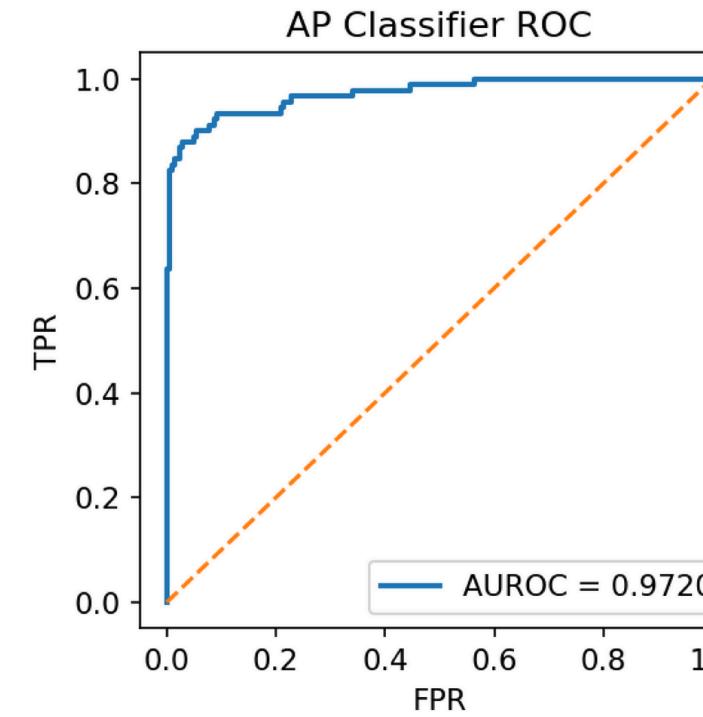


LAT Ensemble



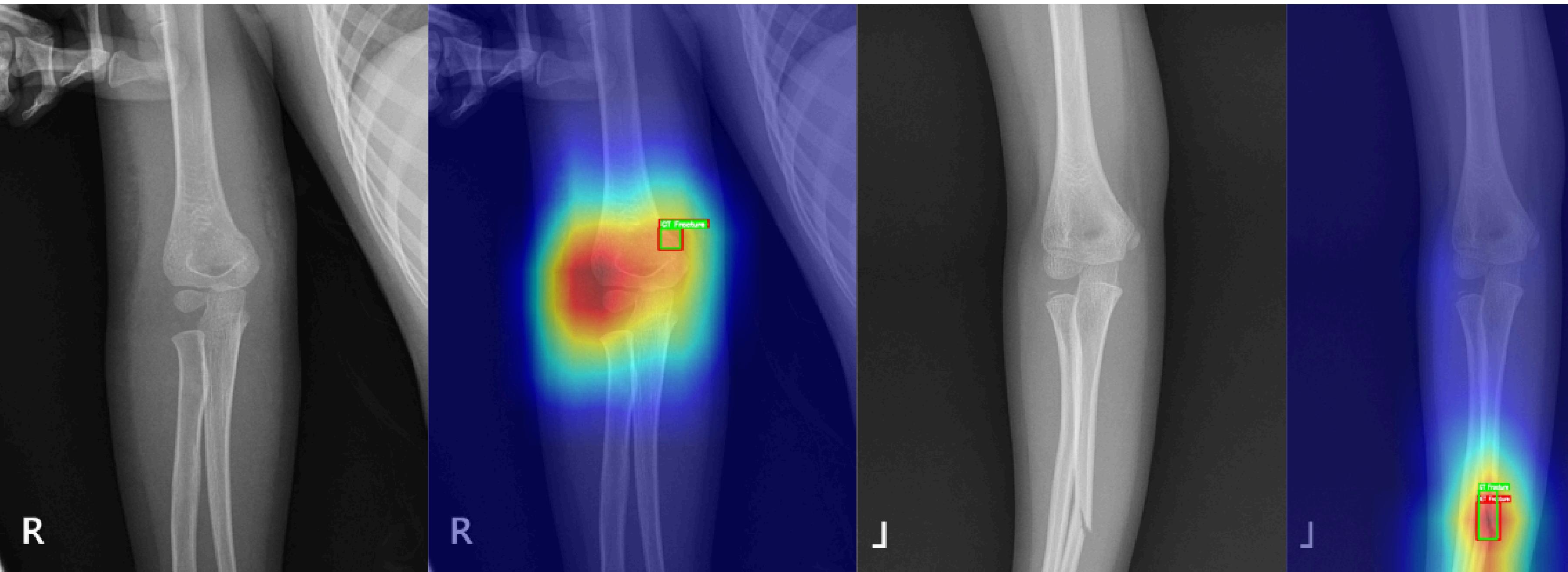
# Ensemble Model

## 2. Result



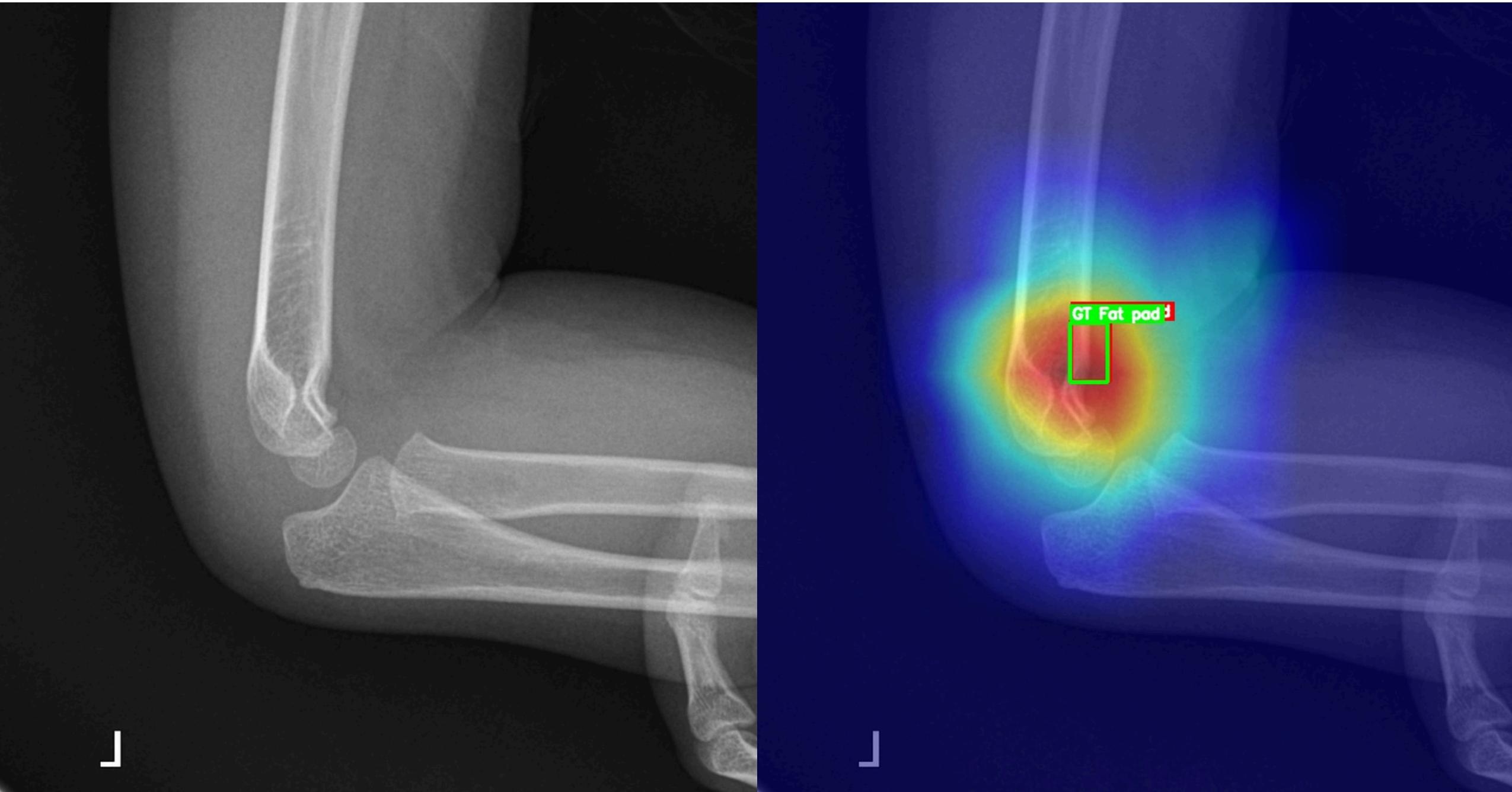
# Ensemble Model

## 2. Result



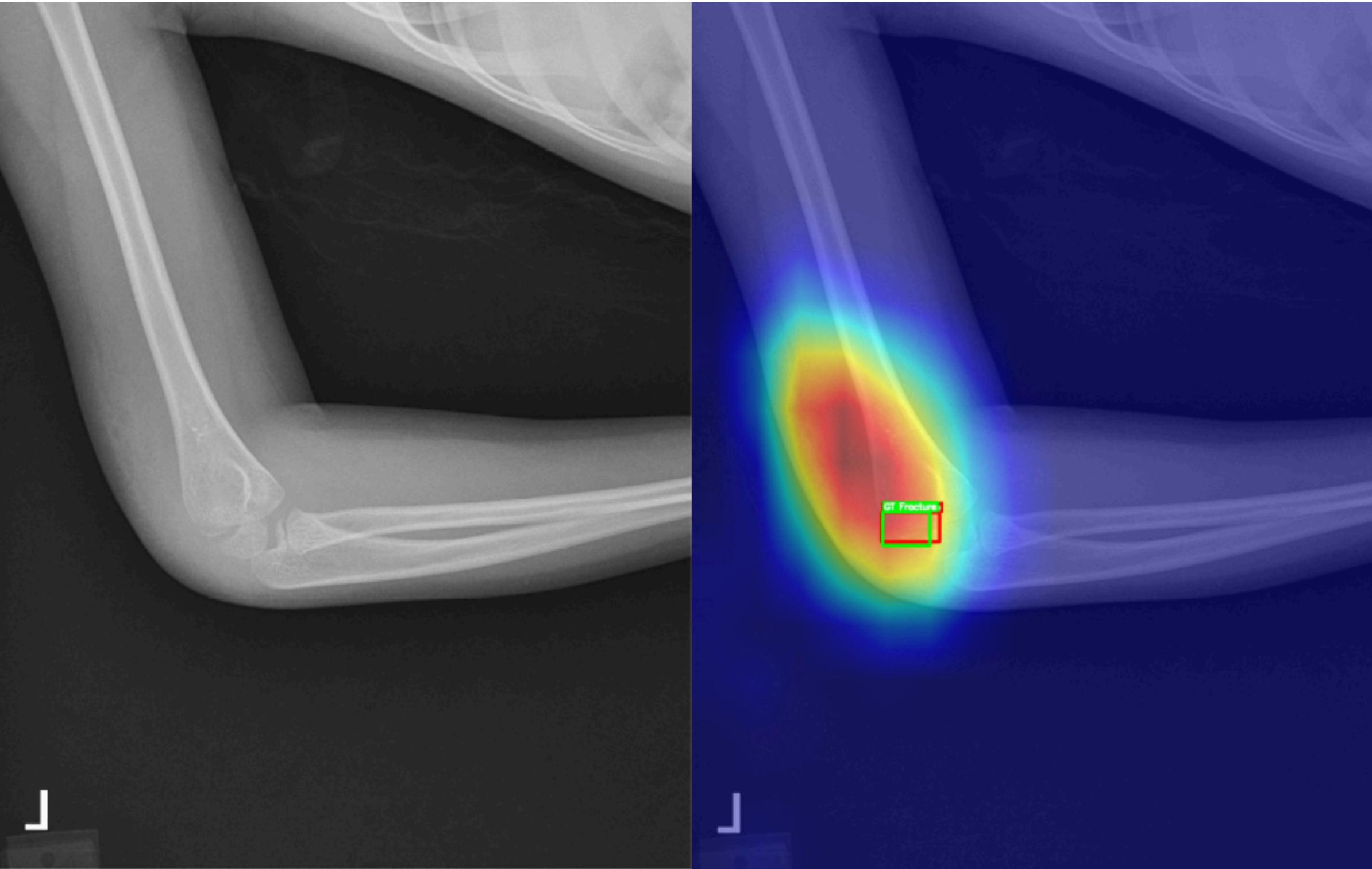
## Ensemble Model

### 2. Result



## Ensemble Model

### 2. Result



---

## External Data

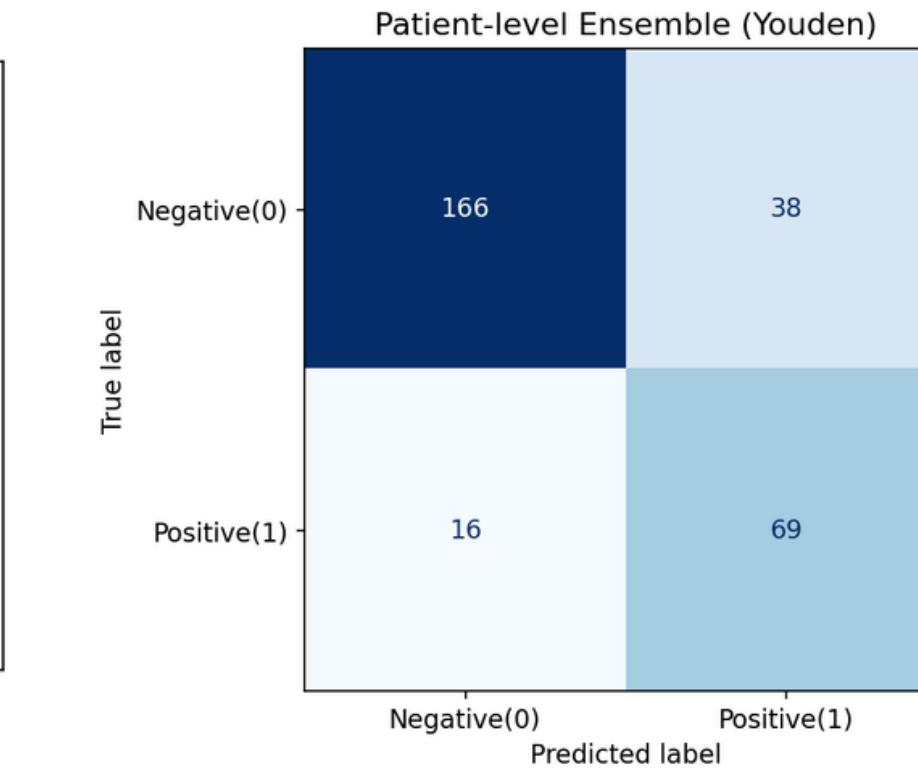
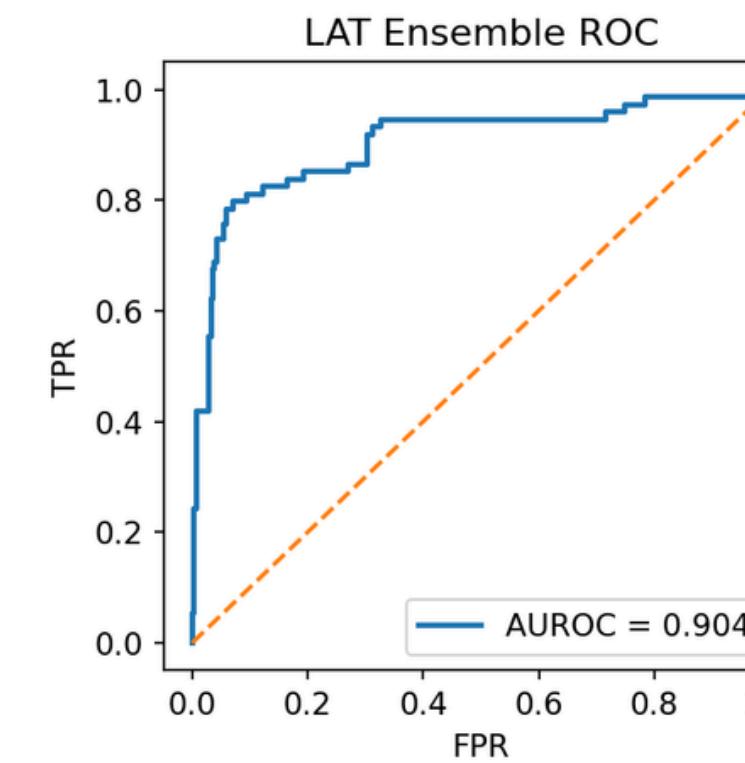
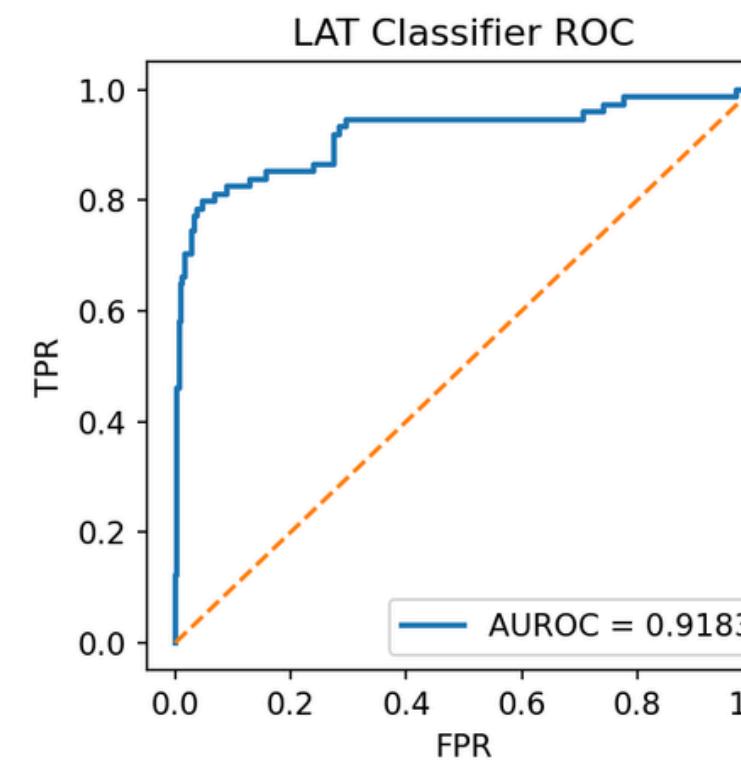
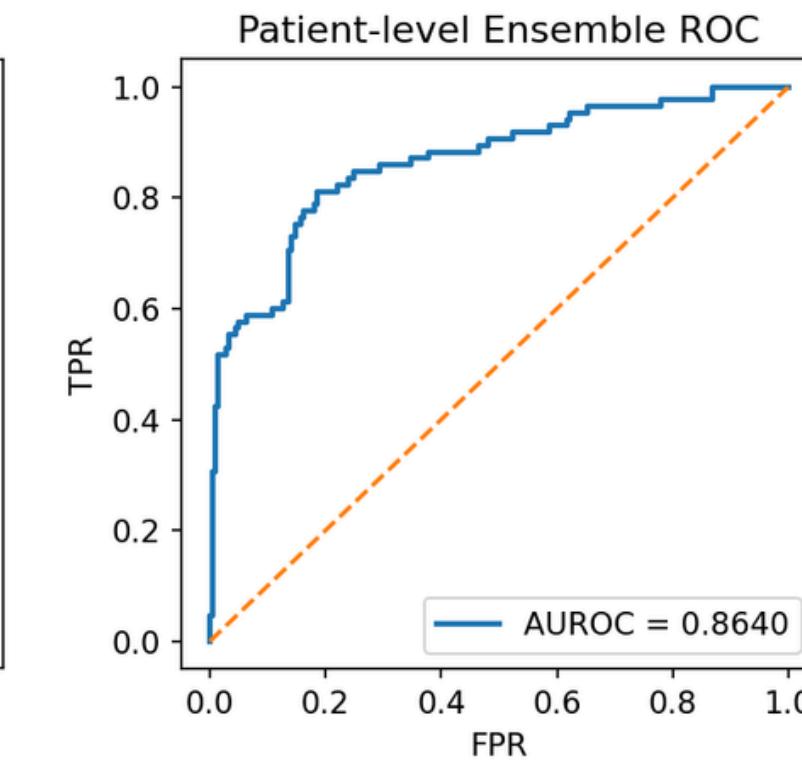
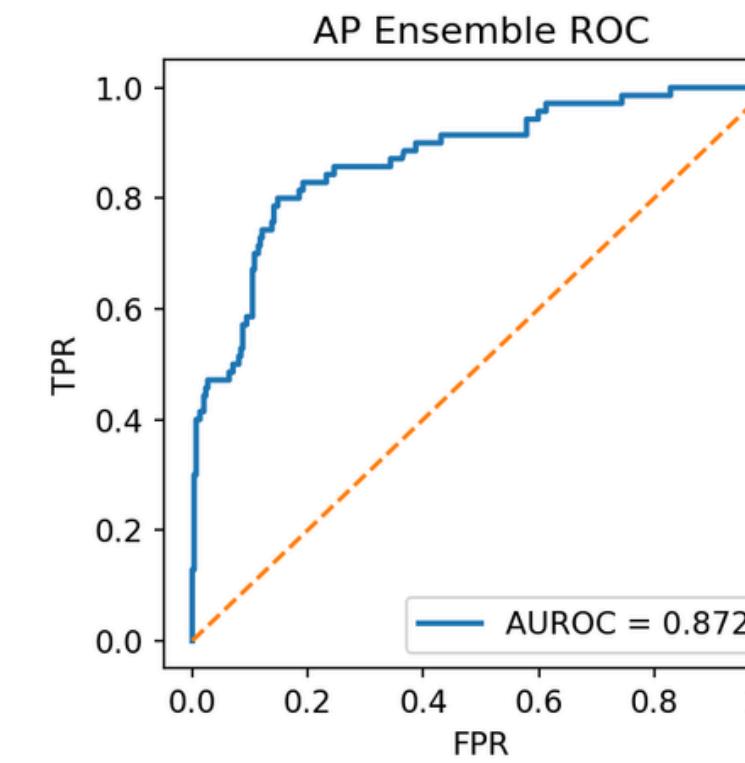
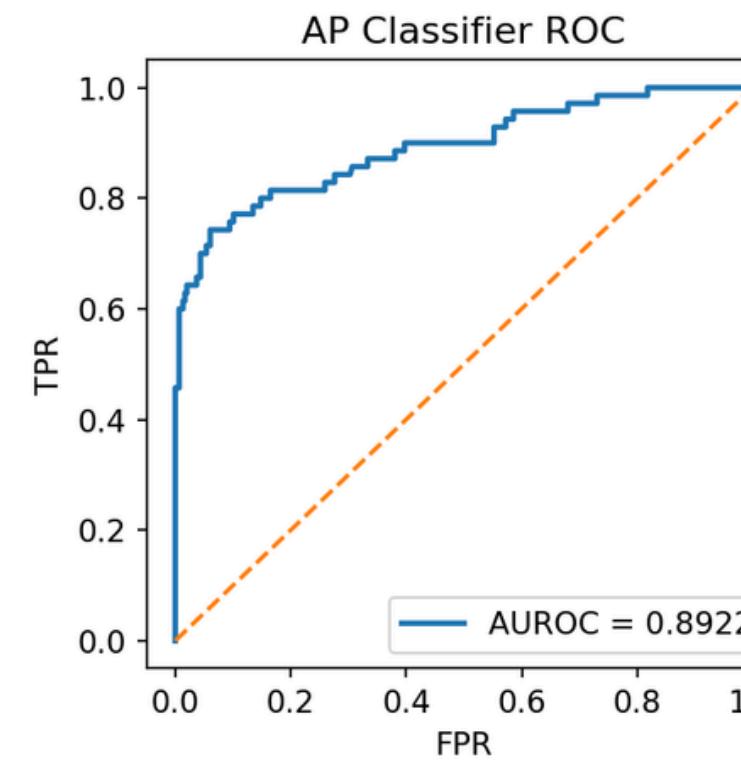
### 1. Dataset

- DICOM 2579개 → PNG 변환시 Error 제외 (Fractures 353, Normal 1767, 미확인 90) → PNG Nosie 제외 (754개)
- 최종 외부검증 Dataset 구성

- AP/Fracture: 69/69
- AP/Normal: 285/285
- AP/Else: 15/15
- Lateral/Fracture: 73/73
- Lateral/Normal: 293/293
- Lateral/Else: 19/19

# External Data

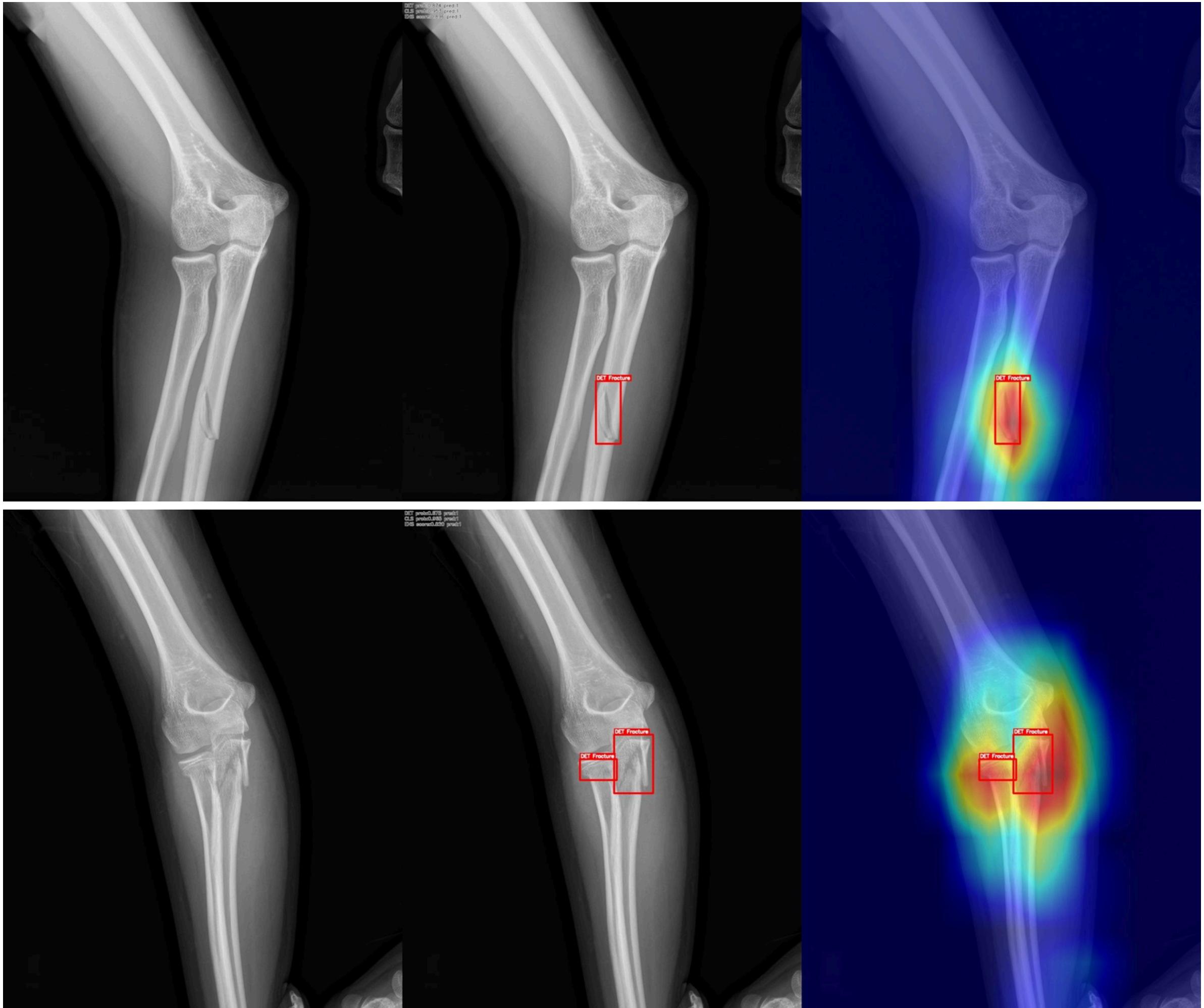
## 1. Result



기술 사항

## External Data

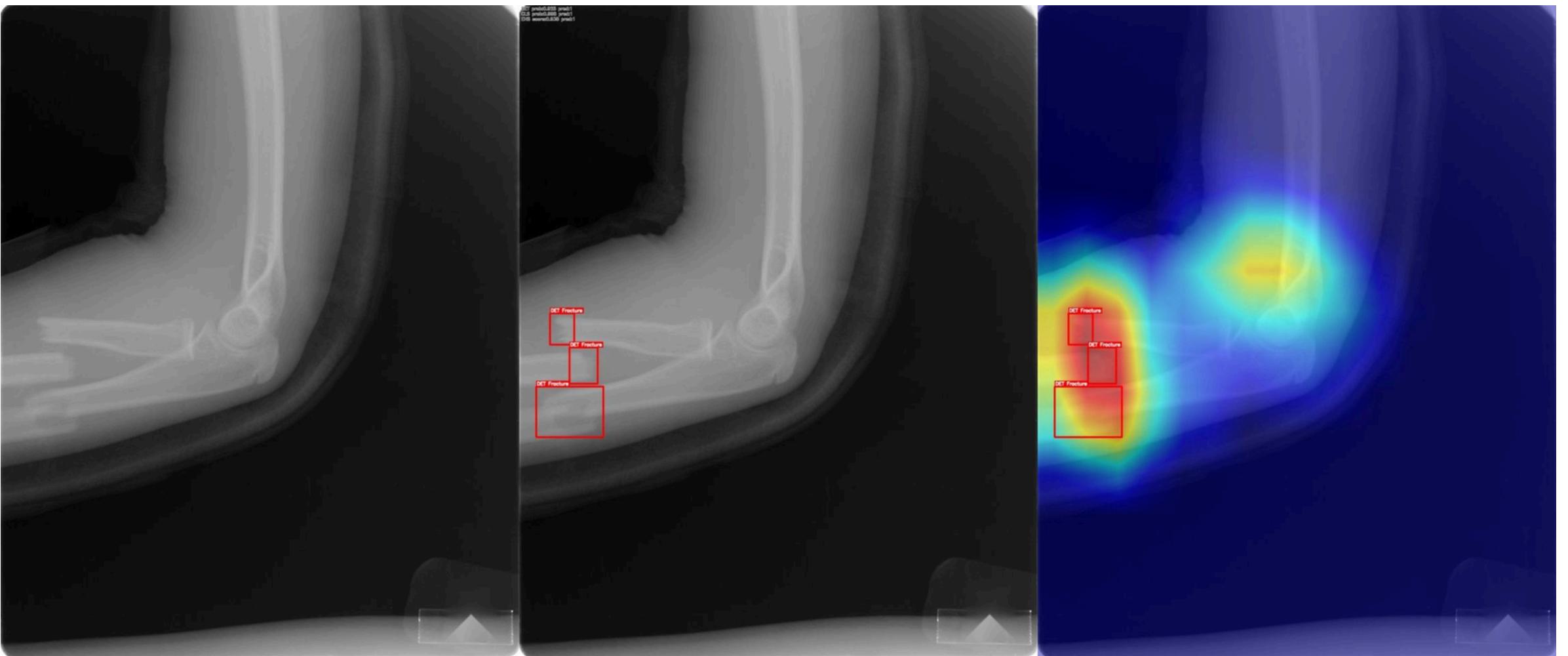
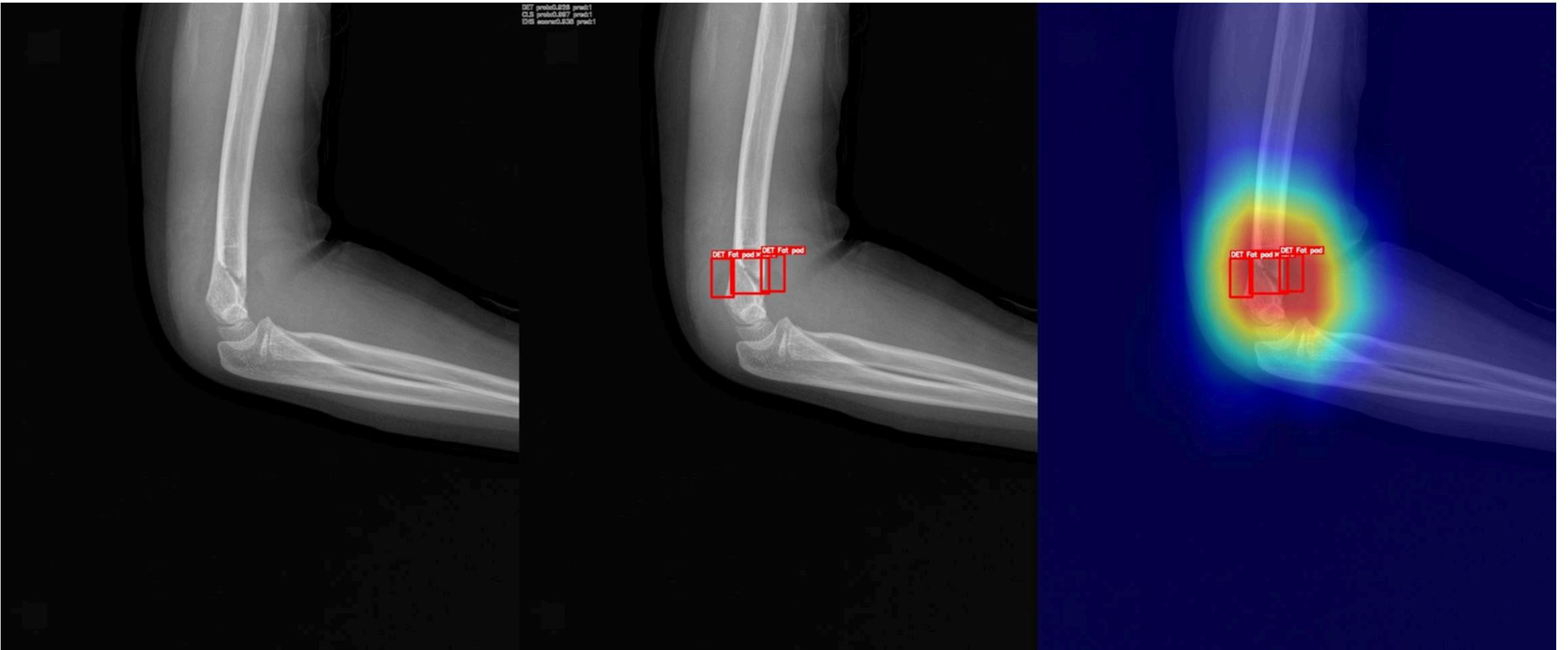
### 1. Result



기술 사항

## External Data

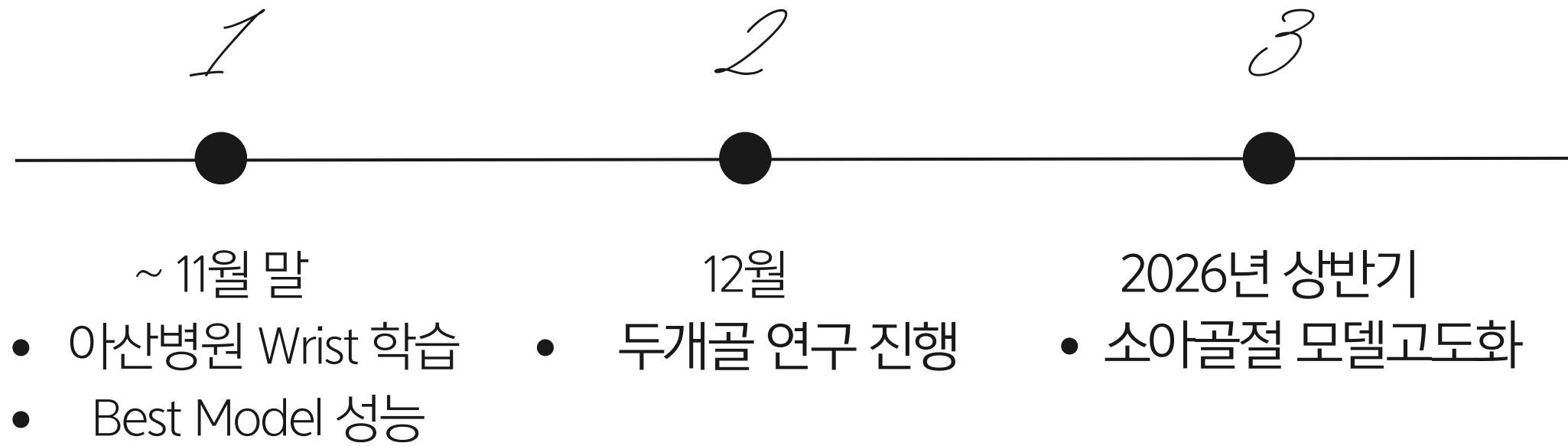
### 1. Result



# 계획



# 진행계획



*Q&A*

# 질의 응답

